

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problem Mailbox.**



## BUNDESREPUBLIK DEUTSCHLAND

PRIORITY  
DOCUMENTSUBMITTED OR TRANSMITTED IN  
COMPLIANCE WITH RULE 17.1(a) OR (b)

DE 99/01867



09/720215

REC L	17 SEP 1999
WIPO	PCT

## Bescheinigung

Die Anmelderin Deutsches Krebsforschungszentrum Stiftung des öffentlichen Rechts in Heidelberg, Neckar/Deutschland hat eine Patentanmeldung unter der Bezeichnung

"Modular aufgebaute RNA-Moleküle mit zwei Sequenzbereichstypen"

am 26. Juni 1998 beim Deutschen Patent- und Markenamt eingereicht.

Die angehefteten Stücke sind eine richtige und genaue Wiedergabe der ursprünglichen Unterlagen dieser Patentanmeldung.

Die Anmeldung hat im Deutschen Patent- und Markenamt vorläufig die Symbole C 12 N, C 07 H und C 07 K der Internationalen Patentklassifikation erhalten.

München, den 30. August 1999

Deutsches Patent- und Markenamt

Der Präsident

Im Auftrag

Aktenzeichen: 198 28 624.4

Dzierzon

Anmelder: Deutsches Krebsforschungszentrum  
Unser Zeichen: K 2529 - hu / msl

### **Modular aufgebaute RNA-Moleküle mit zwei Sequenzbereichstypen**

Die vorliegende Erfindung betrifft RNA-Moleküle, die durch zwei Sequenzbereichstypen gekennzeichnet sind, einen ersten Sequenzbereichstyp, der zur Aufrechterhaltung der dreidimensionalen Struktur des RNA-Moleküls beiträgt, und einen zweiten Sequenzbereichstyp, der für die spezifische Bindung eines Liganden verantwortlich ist. Vorzugsweise sind diese RNA-Moleküle zur direkten Kontrolle der Genexpression von Nutzen. Die vorliegende Erfindung stellt ferner die erfundungsgemäßen RNA-Moleküle enthaltende Vektoren bereit. Ferner betrifft die vorliegende Erfindung Arzneimittel und diagnostische Zusammensetzungen, die die vorstehenden RNA-Moleküle bzw. Vektoren enthalten, einen diese RNA-Moleküle binden-spezifisch erkennenden Antikörper bzw. spezifisch an diese RNA-Moleküle bindende Antisense-RNA oder diese RNA-Moleküle spaltende Ribozyme. Außerdem betrifft die Erfindung nicht-menschliche transgene Säuger und daraus erhaltene Zellen.

Die Regulation der Genexpression bei Eukaryonten erfolgt in der Regel über Proteine, die üblicherweise an bestimmte regulatorische Sequenzen stromaufwärts des zu exprimierenden Gens spezifisch binden und eine charakteristische Wirkung zeigen (RNA-Polymerasen, Transkriptionsfaktoren, Hormon-aktivierbare Rezeptoren etc.). Bisher sind nur wenige Beispiele für die Kontrolle der Genexpression direkt über RNA-Moleküle bekannt. Dazu zählen die für die Inaktivierung des gesamten X-Chromosoms verantwortliche RNA "XIST" ("X-Chromosome inactivation specific transcript"), eine mit IPW bezeichnete RNA ("imprinted in Prader-Willi syndrome") und die RNA H19, die einen Tumorsuppressor darstellt und an der Steuerung bestimmter Entwicklungsvorgänge beteiligt ist. Die artifizielle Kontrolle der Genexpression wird inzwischen durch die Verwendung von spezifisch an mRNAs bindende Antisense-RNAs bewirkt bzw. durch die Verwendung katalytisch aktiver RNA-Moleküle, sogenannter Ribozyme, die an die Ziel-RNA nicht nur spezifisch

binden, sondern diese auch spalten und somit inaktivieren. Allerdings sind die Einsatzmöglichkeiten für diese Antisense-RNAs bzw. Ribozyme begrenzt, vor allem hinsichtlich des zu bindenden und inaktivierenden Liganden, bei dem es sich grundsätzlich nur um RNA handeln kann.

Somit besteht ein Bedarf nach der Bereitstellung von Verbindungen, die universell unterschiedlichste Zielmoleküle, beispielsweise DNA, RNA, Proteine oder niedermolekulare Substanzen, erkennen bzw. inaktivieren können und beispielsweise für die Kontrolle der Genexpression und damit natürlich auch die Prävention und Behandlung von Krankheiten, die mit einer Störung der Genexpression einhergehen, geeignet sind.

Somit liegt der Erfindung im wesentlichen das technische Problem zugrunde, solche Verbindungen bereitzustellen, die unter anderem für die Prävention oder Therapie (und auch Diagnose) von solchen Erkrankungen von Nutzen sind.

Die Lösung dieses technischen Problems wurde durch die Bereitstellung der in den Patentansprüchen gekennzeichneten Ausführungsformen erreicht.

Von den Erfindern konnte ein RNA-Molekül identifiziert werden, das die vorstehend beschriebenen, erwünschten Eigenschaften aufweist. Dieses RNA-Molekül wird durch das Gen "NINTROX" (No INTROns X-chromosome) kodiert, das keine Introns besitzt, auf dem X-Chromosom lokalisiert ist und für kein Protein kodiert. Die genomische Sequenz des menschlichen NINTROX-Gens einschließlich des vermutlichen Promotors ist in Figur 1 gezeigt und die genomische Sequenz des murinen NINTROX-Gens einschließlich des vermutlichen Promotors in Figur 2. In Figur 3 wurde ein Sequenzvergleich zwischen humaner und muriner Sequenz durchgeführt. Daraus sieht man, daß es einige sehr sequenzkonservierte Bereiche gibt, die sich gemäß einer per Computer durchgeführten Energieanalyse durch eine hohe Energie auszeichnen (vgl. Fig. 4).

Während der Wirkungsmechanismus der vorstehend diskutierten, auf RNA-Ebene

wirksamen Gene vollkommen unklar war, konnte durch die Analyse des NINTROX-Gens zum ersten Mal das Wirkprinzip eines solchen Gens, das nachstehend genauer beschrieben wird, ermittelt werden. Das NINTROX-Gen trägt wesentlich zur Aufrechterhaltung der Funktionen des ZNS insbesondere des Hippocampus bei. Defekte in diesem Gen führen zu Einschränkungen der ZNS-Funktionen, die bis zu mentalen Retardierungen reichen. Des Weiteren übt das NINTROX-Gen eine wichtige Funktion bei der Kontrolle der Zellproliferation aus. Dabei können Veränderungen in diesem Gen zu Fehlern in der Kontrolle des Zellwachstums, beispielsweise zu Krebs, führen. Die weiteren Untersuchungen ergaben, daß das Expressionsmuster des NINTROX-Gens gewebe- und entwicklungsspezifisch erfolgt. Die Northern-Analysen zeigten eine Expression in allen untersuchten fotalen und adulten Geweben. Es konnten keine Sequenzhomologien mit bereits bekannten Sequenzen festgestellt werden.

Die Strategie, die zur Identifizierung dieses Nucleinsäuremoleküls führte, wird nachstehend beschrieben. Im Rahmen der systematischen Analyse der q28-Region des menschlichen X-Chromosoms konnten verschiedene exprimierte Sequenzen nachgewiesen und isoliert werden. Mit Hilfe dieser exprimierten Sequenzen konnten einige bisher noch nicht bekannte Gene gemäß Standardmethoden identifiziert und charakterisiert werden, u.a. das der vorliegenden Anmeldung zugrunde liegende NINTROX-Gen. Nachdem das komplette Transkript dieses Gens isoliert war, konnten weder längere offene Leseraster innerhalb des Transkripts noch Introns im Gen nachgewiesen werden. Mit Hilfe der menschlichen Klonen konnte das murine Homolog dieses Gens gemäß Standardmethoden isoliert werden. Auch dieses Gen zeigte keinen offenen Leserahmen oder Introns.

Interessanterweise zeigen die erfindungsgemäßen NINTROX-RNA-Moleküle einen modularen Aufbau, d.h., sie sind durch das Vorhandensein von zwei verschiedenen Sequenzbereichstypen gekennzeichnet. Während der eine Sequenzbereich die Aufrechterhaltung der dreidimensionalen Struktur erlaubt und, wie sich durch Vergleich der Sequenzen aus verschiedenen Spezies (Mensch, Hamster, Känguru, Makaken, Orang-Utan, Schimpansen und Ratte; vgl. Fig. 5) ergibt, nur bedingt

konserviert ist, ist der zweite Sequenzbereich, der für die spezifische Bindung an das Zielmolekül verantwortlich ist, sequenzkonserviert. Aufgrund dieser modularen Bauweise der NINTROX-RNA ist es möglich, diese so zu modifizieren, daß ihre Wirkung nicht nur auf die vorstehend beschriebene Kontrolle der Genexpression beschränkt ist, sondern für eine Vielzahl von Möglichkeiten eingesetzt werden kann. Neben der Kontrolle der Genexpression kann mit Hilfe solcher modular aufgebauter RNA-Moleküle auch die Struktur (z.B. Chromatinstruktur, Nuclear-Scaffold) von chromosomalnen Bereichen verändert werden. Hierdurch ergibt sich die bisher nicht bekannte Möglichkeit, die Expression von größeren genomischen Bereichen gezielt beeinflussen zu können. So können bestimmte Sequenzbereiche der beiden Module des NINTROX-Gens durch andere Sequenzen oder sogar artifizielle Sequenzen ersetzt werden, wodurch (a) die Wechselwirkung dieser RNA mit anderen Bindungspartnern (RNA, DNA, andere Makromoleküle und niedermolekulare Verbindungen) oder deren biochemische Umsetzung (z.B. Erhöhung oder Erniedrigung der Umsatzrate) gezielt verändert werden, wodurch das RNA-Molekül gezielt neuen Aufgaben angepaßt werden kann, und/oder (b) die dreidimensionale Struktur der NINTROX-RNA gezielt spezifischen Anforderungen angepaßt werden kann. Dadurch kann eine teilweise oder völlig neuartige Funktion des erfindungsgemäßen NINTROX-RNA-Moleküls erzielt werden.

Somit betrifft eine Ausführungsform der vorliegenden Erfindung ein RNA-Molekül, das an einen Liganden binden kann und folgende Sequenzbereiche umfaßt: (a) einen die dreidimensionale Struktur des RNA-Moleküls aufrechterhaltenden Sequenzbereich, und (b) einen Sequenzbereich für die spezifische Bindung des Liganden.

Der hier verwendete Ausdruck "einen die dreidimensionale Struktur des RNA-Moleküls aufrechterhaltenden Sequenzbereich" besitzt folgenden Begriffinhalt. Dreidimensionale RNA-Strukturen werden durch Basenpaarung verschiedener Basen innerhalb des RNA-Moleküls ermöglicht. Hierbei werden Strukturen wie "Stems" oder "Loops" gebildet. Viele dieser Strukturen ergeben so die Gesamtstruktur des RNA-Moleküls. Eine Sequenzänderung innerhalb des RNA-Moleküls

kann ohne Folgen für die räumliche Struktur bleiben, wenn die Sequenzveränderung die Basenpaarungen nicht verändert oder wenn die Sequenzänderung durch eine zweite Sequenzänderung ausgeglichen wird. Wird beispielsweise die Basenpaarung A-T zerstört, indem das A zum G mutiert, so kann diese Mutation durch die weitere Mutation des T zum C ausgeglichen werden. Dadurch ändert sich zwar die Sequenz, aber die räumliche Struktur bleibt gleich. Dies hat zur Folge, daß die dieselbe RNA-Struktur durch extrem viele unterschiedliche RNA-Sequenzen gebildet werden kann. Hinweise auf bestimmte RNA-Strukturen ergibt die Analyse der darin enthaltenen Energie. Diese Analyse kann mittels kommerziell erhältlicher Computerprogramme (z.B. "FOLD"; Michael Zuker und P. Stiegler: Optimal Computer Folding of Large RNA Sequences using Thermodynamics and Auxiliary Information, Nucleic Acids Research (81), 9(1), S. 133) durchgeführt werden. Je geringer der Energieinhalt einer bestimmten Sequenz ist, desto stabiler sind die dreidimensionalen RNA-Strukturen. Die Analyse des NINTROX-Gens zeigte eine konservierte Verteilung dieser energiearmen Strukturen (vgl. Fig. 5). Die Basensequenz dieser RNA-Bereiche sind bei verschiedenen Spezies sehr unterschiedlich, der Energieinhalt ist aber äußerst konserviert. In Fig. 3 sind dies die Sequenzbereiche, die nicht durch einen schwarzen Balken am Rand gekennzeichnet sind. Dies bedeutet, daß der die dreidimensionale Struktur des RNA-Moleküls aufrechterhaltenden Sequenzbereich nicht sequenz-, aber energiekonserviert ist. So richten sich Modifikationen dieses Sequenzbereichs auch nicht nach der Basensequenz, sondern nach der Erhaltung des ermittelten Energieinhalts.

Der hier verwendete Ausdruck "ein Sequenzbereich für die spezifische Bindung des Liganden" betrifft einen Sequenzbereich, der so beschaffen ist, daß er den gewünschten Liganden spezifisch binden kann. Diese Sequenzbereiche sind sehr sequenzkonserviert. In Fig. 3 sind diese Bereiche durch einen schwarzen Balken am Rand gekennzeichnet und haben einen hohen Energieinhalt (vgl. Fig. 5). Dies deckt sich mit der Beobachtung, daß diese Sequenzbereiche nicht "verpackt" sind, sondern nach außen gerichtet sind und für die Bindung des Liganden, enzymatische Reaktionen oder die Bindung an andere RNA- oder DNA-Sequenzen verantwortlich sind. Falls es sich bei dem zu bindenden Liganden um ein RNA-Molekül

oder ein DNA-Molekül handelt, wird dieser Sequenzbereich zu einem entsprechenden, ausreichend langen Abschnitt des RNA-Moleküls oder DNA-Moleküls Komplementarität aufweisen. Falls es sich bei dem zu bindenden Liganden um ein Protein handelt, kann der Sequenzbereich (b) ganz oder teilweise durch eine DNA-Sequenz ausgetauscht oder ergänzt werden, von der bekannt ist, daß sie das gewünschte Protein spezifisch bindet.

Die beiden oben beschriebenen Sequenztypen kommen mehrere Male innerhalb der NINTROX-RNA vor. Der Austausch oder die Veränderung einzelner solcher Module ermöglicht die gezielte Veränderung der NINTROX-RNA. So ist bei einer Modifikation des die dreidimensionale Struktur aufrechterhaltenden Moduls auf den ermittelten Energieinhalt zu achten, sodaß dieser einen minimalen Wert beibehält. Die Modifikation des anderen Sequenzbereichs, obwohl dieser gerade als sequenzkonserviert gilt, unterliegt nur geringen Beschränkungen. So kann dieser Bereich ganz oder teilweise weggelassen werden oder kann Insertionen enthalten. Beispielsweise können auch Sequenzen in das NINTROX-RNA-Molekül integriert werden, die bekannte biochemische Eigenschaften aufweisen oder bestimmte DNA-, RNA-Moleküle oder Proteine binden. Darüber hinaus können Zufallssequenzen unterschiedlicher Länge an verschiedenen Stellen des NINTROX-Gens eingebaut werden und danach kann auf spezifische Eigenschaften wie biochemische Umsetzung, spezifische Bindung usw. selektioniert werden.

In einer bevorzugten Ausführungsform des erfindungsgemäßen RNA-Moleküls umfaßt der Sequenzbereich (a) die in Figur 3 nicht am Rand gekennzeichneten Sequenzbereiche oder dazu verwandte Sequenzen, die ebenfalls die Aufrechterhaltung der dreidimensionalen Struktur des RNA-Moleküls erlauben und von dem Sequenzbereich (a) in Figur 3 abweicht. Diese Abweichungen betreffen die Addition, Deletion und/oder Insertion von Basen, wobei der für die Sequenz von Figur (3) ermittelte Energieinhalt zu mindestens 80 %, vorzugsweise mindestens 85 % und mehr bevorzugt zu mindestens 90 % erhalten bleibt. Vorzugsweise bleibt bei diesen eingeführten Änderungen die ursprüngliche dreidimensionale Struktur erhalten.

In einer besonders bevorzugten Ausführungsform umfaßt der Sequenzbereich (b) des erfindungsgemäßen RNA-Moleküls die in Figur 3 dargestellten Sequenzen, die am Rand mit schwarzen Balken versehen sind.

In einer weiteren bevorzugten Ausführungsform des erfindungsgemäßen RNA-Moleküls handelt es sich bei dem zu bindenden Liganden um ein DNA-Molekül oder ein Protein oder Enzym, z.B. DNA-Polymerase I. Vorzugsweise enthält das erfindungsgemäße RNA-Molekül eine Poly(A)-Sequenz am 3'-Ende, was zur Stabilität in einer gewünschten Wirtszelle beitragen kann.

In einer weiteren bevorzugten Ausführungsform wird das erfindungsgemäße RNA-Molekül zur Kontrolle der Genexpression verwendet. Dazu wird der Sequenzbereich (b) so modifiziert, daß er ein für die Genexpression verantwortliches Protein bindet, oder an einen bestimmten DNA-Bereich des Zielgens, wodurch beispielsweise die Anlagerung von Proteinen, die einen die Genexpression hemmenden oder fördern den Einfluß haben, behindert oder unterbunden wird, oder auch direkt an die mRNA des Zielgens, wodurch beispielsweise die Translation behindert oder unterbunden wird. Der Fachmann kann ohne weiteres durch entsprechende Modifikationen des Sequenzbereichs (b) und evtl. auch des Sequenzbereichs (a) das erfindungsgemäße RNA-Molekül so modifizieren, daß es den gewünschten Liganden bindet und so die Genexpression in dem gewünschten Maß kontrolliert.

Die vorliegende Erfindung betrifft auch eine, das erfindungsgemäße RNA-Molekül kodierende DNA-Sequenz, sowie ein Gen mit folgenden Merkmalen: Es enthält einen Promotor, der die Transkription in einer gewünschten Wirtszelle erlaubt sowie eine damit funktionell verknüpfte, das erfindungsgemäße RNA-Molekül kodierende DNA-Sequenz. Vorzugsweise enthält das Gen zusätzlich ein Terminationssignal und eine Polyadenylierungsstelle.

In einer bevorzugten Ausführungsform umfaßt das erfindungsgemäße Gen die in Figur 1 oder 2 dargestellte Sequenz.

Die das erfindungsgemäße RNA-Molekül kodierenden DNA-Sequenzen oder Gene können auch in einen Vektor inseriert werden. Somit umfaßt die vorliegende Erfindung auch diese DNA-Sequenzen oder Gene enthaltende Vektoren. Die Bezeichnung "Vektor" bezieht sich auf ein Plasmid (z.B. pUC18, pBR322, pBlueScript), auf ein Virus oder ein anderes geeignetes Vehikel. In einer bevorzugten Ausführungsform ist die das erfindungsgemäße DNA-Molekül kodierende Sequenz im Vektor mit regulatorischen Elementen funktionell verknüpft, die dessen Expression in prokaryotischen oder eukaryotischen Wirtszellen erlauben. Solche Vektoren enthalten neben den regulatorischen Elementen, beispielsweise einem Promotor, typischerweise einen Replikationsursprung und spezifische Gene, die die phänotypische Selektion einer transformierten Wirtszelle erlauben. Zu den regulatorischen Elementen für die Expression in Prokaryonten, beispielsweise *E.coli*, zählen der lac-, trp-Promotor oder T7-Promotor, und für die Expression in Eukaryonten der AOX1- oder GAL1-Promotor in Hefe, und der CMV-, SV40-, RVS-40-Promotor, CMV- oder SV40-Enhancer für die Expression in tierischen Zellen. Weitere Beispiele für geeignete Promotoren sind der Metallothionein I- und der Polyhedrin-Promotor. Zu geeigneten Vektoren zählen beispielsweise auf T7 basierende Expressionsvektoren für die Expression in Bakterien (Rosenberg et al., Gene 56(1987), 125), pMSXND für die Expression in Säugerzellen (Lee und Nathans, J.Biol.Chem. 263(1988), 3521) und von Baculovirus abgeleitete Vektoren für die Expression in Insektenzellen.

In einer bevorzugten Ausführungsform ist der die erfindungsgemäßen RNA-Moleküle kodierenden Sequenzen enthaltende Vektor ein viraler Vektor, beispielsweise ein *Vaccinia*-Virus oder *Adenovirus*, der bei einer Gentherapie von Nutzen ist. Besonders bevorzugt sind RNA-Viren, vor allem Retroviren. Beispiele für geeignete Retroviren sind MoMuLV, HaMuSV, MuMTV, RSV oder GaLV. Für Zwecke der Gentherapie können die erfindungsgemäßen RNA-Moleküle auch in Form von kolloidalen Dispersionen zu den Zielzellen transportiert werden. Dazu zählen beispielsweise Liposomen oder Lipoplexe (Mannino et al., Biotechniques 6 (1988), 682).

Allgemeine, auf dem Fachgebiet bekannte Verfahren können zur Konstruktion von Expressionsvektoren, die die erfindungsgemäßen RNA-Moleküle kodierenden Sequenzen und geeignete Kontrollsequenzen enthalten, verwendet werden. Zu diesen Verfahren zählen beispielsweise in vitro-Rekombinationstechniken, synthetische Verfahren, sowie in vivo-Rekombinationsverfahren, wie sie beispielsweise in Sambrook et al., beschrieben sind.

Die vorliegende Erfindung betrifft auch die vorstehend beschriebenen Vektoren enthaltende Wirtszellen. Zu diesen Wirtszellen zählen Bakterien, Hefe, Insekten- und Tierzellen, vorzugsweise Säugerzellen. Bevorzugte Säugerzellen sind CHO-, VERO-, BHK-, HeLa-, COS-, MDCK, 293- und WI38-Zellen. Verfahren zur Transformation dieser Wirtszellen, zur phänotypischen Selektion von Transformanten und zur Expression der erfindungsgemäßen Nucleinsäuremoleküle unter Verwendung der vorstehend beschriebenen Vektoren sind auf dem Fachgebiet bekannt.

Die vorliegende Erfindung betrifft auch Antikörper, die das erfindungsgemäße RNA-Molekül spezifisch erkennen. Die Antikörper können monoklonale, polyclonale oder synthetische Antikörper sein oder Fragmente davon, beispielsweise Fab-, Fv- oder scFv-Fragmente. Vorzugsweise handelt es sich dabei um einen monoklonalen Antikörper. Die erfindungsgemäßen Antikörper können gemäß Standardverfahren hergestellt werden, wobei das erfindungsgemäße RNA-Molekül oder ein Fragment davon als Immunogen dienen. Monoklonale Antikörper können beispielsweise durch das von Köhler und Milstein (Nature 256 (1975), 495) und Galfré (Meth. Enzymol. 73 (1981), 3) beschriebene Verfahren hergestellt werden, wobei Maus-Myelomzellen mit von immunisierten Säugern stammenden Milzzellen fusioniert werden. Diese Antikörper können beispielsweise zur Hemmung der Aktivität der erfindungsgemäßen RNA-Moleküle, beispielsweise zur Beeinflussung der Genexpression, verwendet werden. Die Antikörper können beispielsweise auch in diagnostischen Assays verwendet werden, um nachweisen zu können, ob eine Dysregulation der Genexpression, beispielsweise durch einen Verlust oder Mangel der verantwortlichen NINTROX-RNA einhergeht. Die Antikörper können in Immu-

nassays in Flüssigphase vorliegen oder an einen festen Träger gebunden werden. Dabei können die Antikörper auf verschiedene Art und Weise markiert sein. Geeignete Marker und Markierungsverfahren sind auf dem Fachgebiet bekannt. Beispiele für Immunassays sind ELISA und RIA.

Die Erfindung betrifft ferner Antisense-RNAs, die an ein erfindungsgemäßes RNA-Molekül spezifisch binden und zur Herabsetzung der Expression von Genen, die unter direkter Kontrolle von RNA, beispielsweise NINTROX-RNA, stehen, in vitro oder in vivo verwendet werden können. Die Verabreichung der erfindungsgemäß Antisense-RNA an eine Zielzelle führt zu einer herabgesetzten Genexpression und ist für die Behandlung von Erkrankungen besonders nützlich, die durch eine zu hohe Genexpression des unter direkter RNA-Kontrolle stehenden Gens gekennzeichnet sind (z.B. bei Krebserkrankungen). Dabei können die Antisense-RNAs direkt verabreicht werden oder als diese kodierende DNA, vorzugsweise in einen geeigneten Vektor inseriert. Zu den geeigneten Vektoren zählen alle bereits vorstehend in Zusammenhang mit den erfindungsgemäß RNA-Molekülen beschriebenen Vektoren.

Die erfindungsgemäß Antisense-RNAs umfassen eine Antisense-Sequenz mit mindestens 7 bis 10 oder mehr Nucleotiden, die mit einer Sequenz des erfindungsgemäß RNA-Moleküls, beispielsweise NINTROX-RNA, spezifisch hybridisieren. Vorzugsweise weist die erfindungsgemäß Antisense-RNA eine Länge von etwa 10 bis etwa 50 Nucleotiden oder von etwa 14 bis etwa 35 Nucleotiden auf. In weiteren Ausführungsformen handelt es sich bei den erfindungsgemäß Antisense-RNAs um RNAs, die kürzer als etwa 100 Nucleotide oder kürzer als etwa 200 Nucleotide sind. Im allgemeinen sollten die Antisense-RNAs lang genug sein, um eine stabile Doppelhelix zu bilden, jedoch kurz genug (in Abhängigkeit von der Art der Zuführung) um, falls erwünscht, in vivo verabreicht werden zu können. Im allgemeinen ist die Antisense-Sequenz zur Gewährleistung einer spezifischen Hybridisierung zu der Ziel-Sequenz im wesentlichen komplementär. In bestimmten Ausführungsformen ist die Antisense-Sequenz genau komplementär zu der Zielsequenz. Die Antisense-RNAs können jedoch auch Nucleotid-Substitutionen, Ad-

ditionen, Deletionen, Transitionen, Transpositionen oder Modifikationen enthalten, solange die spezifische Bindung an die relevante Zielsequenz als eine funktionelle Eigenschaft der Antisense-RNA beibehalten wird. Die Antisense-RNAs können auch zusätzlich zu den Antisense-Sequenzen weitere Sequenzen enthalten. Die Antisense-RNAs (sowie die erfindungsgemäßen RNA-Moleküle) können unter Verwendung jedes zur Herstellung von Nucleinsäuren geeigneten Verfahrens hergestellt werden, beispielsweise durch chemische Synthese *de novo* oder durch Clonierung. Eine Antisense-RNA kann beispielsweise auch dadurch hergestellt werden, daß eine Sequenz der Ziel-RNA oder eines Fragments davon in umgekehrter Orientierung funktionell mit einem Promotor verknüpft in einen Vektor (z.B. ein Plasmid) inseriert wird. Unter der Voraussetzung, daß der Promotor und vorzugsweise Terminations- und Polyadenylierungssignale korrekt positioniert sind, wird der Strang der inserierten Sequenz, der dem nicht-codierenden Strang entspricht, transkribiert, und dieser wirkt als eine Antisense-RNA.

Die vorliegende Erfindung betrifft auch Ribozyme, die die erfindungsgemäßen RNA-Moleküle spezifisch spalten und somit auch zur Hemmung der Genexpression von Nutzen sind. Nützliche Ribozyme können 5'- und 3'-terminale Sequenzen umfassen, die zu der Ziel-RNA komplementär sind, und diese können vom Fachmann nach Standardverfahren konstruiert werden (siehe beispielsweise PCT-Veröffentlichung WO 93/23572). Zu den erfindungsgemäßen Ribozymen gehören beispielsweise Ribozyme mit den Merkmalen der Gruppe I-Intron-Ribozyme (Cech, Biotechnology 13 (1995), 323) und "hammerhead"-Ribozyme (Edgington, Biotechnology 10 (1992), 256).

In einer Ausführungsform werden die erfindungsgemäßen Ribozyme per se als Arzneimittel verwendet. In einer anderen Ausführungsform werden Gentherapieverfahren zur Expression von Ribozymen in einer Zielzelle *ex vivo* oder *in vivo* angewandt. Die Verfahren zur Verabreichung der Ribozyme bzw. zur Expression der Ribozyme *in vivo* entsprechen den vorstehend in Zusammenhang mit den erfindungsgemäßen RNA-Molekülen beschriebenen Verfahren.

Die Isolierung und Charakterisierung des menschlichen NINTROX-Gens und insbesondere des Maushomologs des NINTROX-Gens erlaubt die Etablierung eines Tiermodells, das die Bereitstellung von Therapien und Arzneimitteln für die vorstehend diskutierten Krankheiten erlaubt. Durch die Bereitstellung der Sequenz des NINTROX-Gens ist sowohl eine Diagnose (post- oder pränatal) als auch eine Therapie von Erkrankungen möglich, bei denen die Genexpression durch das Fehlen von NINTROX-RNA oder einen Überschuß von NINTROX-RNA gekennzeichnet ist. Die therapeutische oder diagnostische Anwendung ist jedoch nicht nur auf Krankheiten beschränkt, die mit einer Fehlregulation der Expression eines Gens einhergehen, das unter Kontrolle der NINTROX-RNA steht, sondern die entsprechend den vorstehend beschriebenen Möglichkeiten veränderten RNA-Moleküle bieten darüber hinaus die Möglichkeit völlig neuer Therapeutika.

Die vorliegende Erfindung betrifft somit ferner Arzneimittel, die die vorstehend beschriebenen RNA-Moleküle, Vektoren, Antikörper, Antisense-RNAs oder Ribozyme enthalten. Diese Arzneimittel enthalten gegebenenfalls zusätzlich einen pharmazeutisch verträglichen Träger. Geeignete Träger und die Formulierung derartiger Arzneimittel sind dem Fachmann bekannt. Zu geeigneten Trägern zählen beispielsweise Phosphat-gepufferte Kochsalzlösungen, Wasser, Emulsionen, beispielsweise Öl/Wasser-Emulsionen, Netzmittel, sterile Lösungen etc. Die Verabreichung der Arzneimittel kann oral oder parenteral erfolgen. Zu den Verfahren für die parenterale Verabreichung gehören die topische, intra-arterielle (z.B. direkt zu einem Tumor), intramuskuläre, subkutane, intramedulläre, intrathekale, intraventrikuläre, intraveneöse, intraperitoneale oder intranasale Verabreichung. Die geeignete Dosierung wird von dem behandelnden Arzt bestimmt und hängt von verschiedenen Faktoren ab, beispielsweise von dem Alter, dem Geschlecht, dem Gewicht des Patienten, dem Stadium eines Tumors, der Art der Verabreichung etc..

Vorzugsweise wird das erfindungsgemäße Arzneimittel zur Prävention oder Behandlung von Erkrankungen verwendet, die mit einer Störung der Kontrolle der Genexpression in Zusammenhang stehen. Besonders bevorzugt wird das erfindungsgemäße Arzneimittel zur Behandlung von Tumorerkrankungen oder Erkrankungen, die mit einer Störung der Genexpression in Zusammenhang stehen.

kungen des ZNS verwendet. Dabei kann das Arzneimittel in der Gentherapie Verwendung finden, wobei die vorstehend beschriebenen Verfahren bzw. Vektoren zur Einschleusung der erfindungsgemäßen Nucleinsäuren Anwendung finden können. Andererseits kann das erfindungsgemäße RNA-Molekül direkt verabreicht werden, um so in Zellen, die keine funktionalen Kopien des RNA-Moleküls mehr besitzen, normale Expression des Gens wiederherzustellen.

Die vorliegende Erfindung betrifft ferner eine diagnostische Zusammensetzung, die das erfindungsgemäße RNA-Molekül, die dieses kodierende DNA-Sequenz oder ein Fragment davon, den erfindungsgemäßen Antikörper oder ein Fragment davon, oder die erfindungsgemäße Antisense-RNA oder ein Fragment davon enthält oder Kombinationen davon, gegebenenfalls zusammen mit einem geeigneten Nachweismittel. Mittels dieser diagnostischen Zusammensetzung kann der Nachweis darüber erfolgen, ob die direkt die Genexpression steuernde RNA, beispielsweise NINTROX-RNA vorhanden ist oder im Vergleich zu einer Kontrolle in zu hoher oder niedriger Konzentration oder mit einer abweichenden Länge vorliegt. Dabei wird vorzugsweise der Antikörper oder ein Fragment davon in den vorstehend beschriebenen Assays oder die Antisense-RNA oder ein Fragment davon als Sonde in Hybridisierungsexperimenten verwendet. Vorzugsweise weist dazu die Sonde eine Länge von mindestens 10, besonders bevorzugt mindestens 15 Basen auf. Geeignete, auf Hybridisierung basierende Nachweisverfahren sind dem Fachmann bekannt. Geeignete Markierungen für die Sonde sind dem Fachmann ebenfalls bekannt und dazu zählen beispielsweise Markierung mit Radioisotopen, Biolumineszenz-, Chemilumineszenz-, Fluoreszenzmarkern, Metallchelaten, Enzymen etc. Bei diesem Verfahren können dem Fachmann bekannte Methoden bezüglich der Präparation von Gesamt-RNA bzw. poly(A) + RNA aus biologischen Proben, der Auf trennung der RNAs auf nach Größe auftrennenden Gelen, beispielsweise denaturierenden Agarosegelen, der Herstellung und Markierung der Sonde und des Nachweises der Hybride, beispielsweise über "Northern-Blot"; angewandt werden. Vorzugsweise erfolgt dabei die Diagnose von Erkrankungen, wie sie vorstehend im Zusammenhang mit den erfindungsgemäßen Arzneimitteln beschrieben wurden.

Eine Diagnose kann auch auf DNA-Ebene erfolgen. Dabei wird mit den vorstehend beschriebenen Nucleinsäuremolekülen die Intaktheit des Gens, das die an der Regulation der Genexpression direkt beteiligte RNA, beispielsweise NINTROX-RNA, kodiert untersucht (beispielsweise hinsichtlich Vorhandensein, Länge oder Mutationen). Bei diesem Verfahren können dem Fachmann bekannte Methoden bezüglich der Präparation von DNA aus biologischen Proben, des Restriktionsverdaus der DNA, der Auftrennung der Restriktionsfragmente auf nach Größe auftrennenden Gelen, beispielsweise Agarosegelen, der Herstellung und Markierung der Sonde und des Nachweises der Hybridisierung, beispielsweise über "Southern-Blot"; angewandt werden. Der vorstehende Nachweis kann auch über PCR durchgeführt werden. Dabei werden Primer verwendet, die die kodierende Sequenz flankieren. Diagnostisch von Bedeutung sind dabei Amplifikationsprodukte von DNA aus dem fraglichen Gewebe, die sich, beispielsweise hinsichtlich ihrer Länge oder Sequenz, von den Amplifikationsprodukten von DNA aus gesundem Gewebe unterscheiden.

Gegenstand der vorliegenden Erfindung ist ferner ein nicht-menschliches Säugetier, dessen NINTROX-Gen verändert ist, z.B. durch Insertion einer heterologen Sequenz, insbesondere einer Selektionsmarkersequenz.

Der Ausdruck "nicht-menschliches Säugetier" umfaßt jegliches Säugetier, dessen GR in seiner induzierenden Funktion verändert sein kann. Beispiele solcher Säugetiere sind Maus, Ratte, Kaninchen, Pferd, Rind, Schaf, Ziege, Affe, Schwein, Hund und Katze, wobei Maus bevorzugt ist.

Der Ausdruck "NINTROX-Gen, das verändert ist" bedeutet, daß in dem im nicht-menschlichen Säugetier natürlich vorkommenden NINTROX-Gen durch Standardmethoden eine Deletion von ca. 1-2 kb durchgeführt und an dessen Stelle eine heterologe Sequenz, z.B. ein Konstrukt zur Vermittlung von Antibiotika-Resistenz (z.B. eine "neo-Kassette") eingefügt wird. Diese Methode ist allgemein in Schwartzberg et al., Proc. Natl. Acad. Sci. USA, Vol. 87, S. 3210-3214, 1990 beschrieben, worauf hier Bezug genommen wird.

Ein weiterer Gegenstand der vorliegenden Erfindung sind Zellen, die aus dem vorstehenden nicht-menschlichen Säugetier erhalten werden. Diese Zellen können in jeglicher Form vorliegen, z.B. in einer Primär- oder Langzeit-Kultur.

Ein erfindungsgemäßes nicht-menschliches Säugetier kann durch übliche Verfahren bereitgestellt werden. Günstig ist ein Verfahren, das folgende Schritte umfaßt:

- (a) Herstellung eines DNA-Fragments, insbesondere eines Vektors, enthaltend ein verändertes NINTROX-Gen, wobei das NINTROX-Gen durch Insertion einer heterologen Sequenz, insbesondere eines selektierbaren Markers, verändert worden ist;
- (b) Präparation embryonaler Stammzellen aus einem nicht-menschlichen Säuger (bevorzugt Maus);
- (c) Transformation der embryonalen Stammzellen von Schritt (b) mit dem DNA-Fragment von Schritt (a), wobei das NINTROX-Gen in den embryonalen Stammzellen durch homologe Rekombination mit dem DNA-Fragment von (a) verändert wird,
- (d) Kultivieren der Zellen von Schritt (c),
- (e) Selektion der kultivierten Zellen von Schritt (d) auf das Vorhandensein der heterologen Sequenz, insbesondere des selektierbaren Markers,
- (f) Erzeugen chimerer nicht-menschlicher Säuger aus den Zellen von Schritt (e) durch Injektion dieser Zellen in Säuger-Blastozysten (bevorzugt Maus-Blastozysten), Übertragen der Blastozysten in pseudo-schwangere weibliche Säuger (bevorzugt Maus) und Analyse der erhaltenen Nachkommen auf eine Veränderung des NINTROX-Gens.

In Schritt (c) wird der Mechanismus der homologen Rekombination (vgl. R.M.

Torres, R. Kühn, Laboratory Protocols for Conditional Gene Targeting, Oxford University Press, 1997) ausgenutzt, um embryonale Stammzellen zu transfizieren. Die homologe Rekombination zwischen den in einem Chromosom vorhandenen DNA-Sequenzen und neuen, hinzugefügten clonierten DNA-Sequenzen ermöglicht das Einfügen eines klonierten Gens in das Genom einer lebenden Zelle anstelle des ursprünglichen Gens. Mit dieser Methode können bei Verwendung embryonaler Keimzellen via Chimären Tiere erhalten werden, die für das gewünschte Gen oder den gewünschten Genteil oder die gewünschte Mutation homozygot sind.

Der Ausdruck "embryonale Stammzellen" betrifft jegliche embryonalen Stammzellen eines nicht-menschlichen Säugetiers, die sich zur Mutierung des NINTROX-Gens eignen. Vorzugsweise sind die embryonalen Stammzellen von der Maus, insbesondere die Zellen E14/1 oder 129/SV.

Der Ausdruck "Vektor" umfaßt jeglichen Vektor, der durch Rekombination mit der DNA von embryonalen Stammzellen eine Veränderung des NINTROX-Gens ermöglicht. Vorzugsweise weist der Vektor einen Marker auf, mit dem auf vorhandene Stammzellen selektiert werden kann, in denen die gewünschte Rekombination erfolgt ist. Ein solcher Marker ist z.B. die loxP/tkneo-Cassette, die mit Hilfe des Cre/loxP-Systems wieder aus dem Genom entfernt werden kann.

Des Weiteren kennt der Fachmann Bedingungen und Materialien, um die Schritte (a)-(f) durchzuführen.

Mit der vorliegenden Erfindung wird ein nicht-menschliches Säugetier bereitgestellt, dessen NINTROX-Gen verändert ist. Diese Veränderung kann ein Ausschalten der Genexpression-regulierenden Funktion sein. Mit einem solchen Säugetier bzw. Zellen daraus kann selektiv die Genexpression-kontrollierende Funktion von NINTROX untersucht werden. Ferner ist es hiermit möglich, Substanzen, Arzneimittel und Therapieansätze zu finden, mit denen selektiv auf die kontrollierende Funktion von NINTROX eingewirkt werden kann. Daher liefert die vorliegende Erfindung eine Basis, um auf die verschiedensten Erkrankungen einzuwirken.

Solche Erkrankungen sind z.B. Einschränkungen der ZNS-Funktionen, die bis zu mentalen Retardierungen reichen oder die Induktion von Krebs durch Fehler bei der Kontrolle der Zellproliferation. Ferner sollte es möglich sein, die Rolle des Hippocampus näher zu untersuchen und zu charakterisieren.

Die folgenden Klone wurden am 4. Mai 1998 bei der DSMZ, Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH, Mascheroder Weg 1b, D-38124 Braunschweig, hinterlegt:

DSM 12153: E.coli JFC-484, Teilsequenz der humanen NINTROX-cDNA  
DSM 12154: E.coli JFC-622, Teilsequenz der murinen NINTROX-cDNA  
DSM 12155: E.coli JFC-8D3, Sequenz der humanen genomischen NINTROX-DNA  
DSM 12156: E.coli JFC-P1-165, Sequenz der murinen genomischen NINTROX-DNA

Die Figuren zeigen:

Figur 1: Humane Sequenz des NINTROX-Gens (einschließlich des vermutlichen Promotors)

Figur 2: Murine Sequenz des NINTROX-Gens (einschließlich des vermutlichen Promotors)

Figur 3: Sequenzvergleich humane (oben) und murine (unten) Sequenz  
gestrichelt: Promotor  
durchgezogener Balken: Sequenzkonservierte Bereiche (b)

Figur 4: Energiediagramm der Sequenzen aus Fig. 3

Figur 5: Homologievergleich von NINTROX aus verschiedenen Spezies

Das folgende Beispiel erläutert die Erfindung:

**Beispiel 1: Identifizierung und Charakterisierung des NINTROX-Gens**

Zur Identifikation von transkribierten Sequenzen aus der Region Xq27.3 bis Xqter wurde zunächst aus verschiedenen Geweben des Schweins (Niere, Herz, Milz, Leber, Gehirn usw.) Gesamt-RNA isoliert und mit Hilfe von Oligo-dT in Erststrang-cDNA überschrieben. Diese komplexen cDNA-Proben, die alle in dem jeweiligen Gewebe transkribierten Gene repräsentieren, wurden dann radioaktiv markiert und mit der Xq27.3-Xqter spezifischen Cosmid-Bibliothek hybridisiert. Die Cosmid-Bibliothek wurde dabei in Form von systematisch auf Nylonmembranen angeordneten Cosmid-Klonen analysiert. Anschließend wurde von den Cosmid-Klonen, die mit den komplexen cDNA-Proben positive Hybridisierungssignale aufwiesen, die Cosmid-DNA isoliert, mit EcoRI verdaut, durch Gel-elektrophorese getrennt und auf Nylonmembranen transferiert. Die Restriktionsfragmente, die dann eine positive Hybridisierung mit den komplexen, radioaktiv markierten cDNA-Proben aufwiesen, wurden dann isoliert und radioaktiv markiert und zum Durchsuchen einer fötalen humanen cDNA-Bank verwendet. Hierdurch konnten positive cDNA-Klone isoliert werden, die das Transkript des NINTROX-Gens repräsentierten.

### Patentansprüche

1. RNA-Molekül, das an einen Liganden binden kann und folgende Sequenzbereiche umfaßt:
  - (a) einen die dreidimensionale Struktur des RNA-Moleküls aufrechterhaltenden Sequenzbereich; und
  - (b) einen Sequenzbereich für die spezifische Bindung des Liganden.
2. RNA-Molekül nach Anspruch 1, wobei der Sequenzbereich (a) die in Figur 3 ohne Balken am Rand dargestellte Sequenz umfaßt oder eine dazu verwandte Sequenz, die ebenfalls die Aufrechterhaltung der dreidimensionalen Struktur des RNA-Moleküls erlaubt.
3. RNA-Molekül nach Anspruch 1 oder 2, wobei der Sequenzbereich (b) die in Figur 3 mit Balken am Rand dargestellte Sequenz umfaßt.
4. RNA-Molekül nach einem der Ansprüche 1 bis 3, wobei der Ligand ein DNA-Molekül oder ein Protein ist.
5. RNA-Molekül nach einem der Ansprüche 1 bis 4, das zusätzlich eine Poly(A)-Sequenz am 3'-Ende enthält.
6. RNA-Molekül nach einem der Ansprüche 1 bis 5 zur Kontrolle der Genexpression.
7. DNA-Sequenz, die ein RNA-Molekül nach einem der Ansprüche 1 bis 6 kodiert.
8. Gen, das die in Figur 1 oder 2 dargestellte Sequenz umfaßt.
9. Vektor, der die DNA-Sequenz nach Anspruch 7 oder das Gen nach An-

spruch 8 enthaltend.

10. Vektor nach Anspruch 9, wobei der Vektor ein Plasmid ist.
11. Vektor nach Anspruch 10, wobei der Vektor ein viraler Vektor ist.
12. Vektor nach Anspruch 11, der ein RNA-Virus ist.
13. Vektor nach Anspruch 12, der ein Retrovirus ist.
14. Wirtszelle, den Vektor nach einem der Ansprüche 9 bis 13 enthaltend.
15. Wirtszelle nach Anspruch 14, wobei die Wirtszelle eine Säugerzelle ist.
16. Antikörper oder ein Fragment davon, die ein RNA-Molekül nach einem der Ansprüche 1 bis 6 spezifisch binden.
17. Antikörper nach Anspruch 16, wobei der Antikörper ein monoklonaler Antikörper ist.
18. Antisense-RNA, die an ein RNA-Molekül nach einem der Ansprüche 1 bis 6 spezifisch bindet.
19. Ribozym, das ein RNA-Molekül nach einem der Ansprüche 1 bis 6 spezifisch spaltet.
20. Verwendung des RNA-Moleküls nach einem der Ansprüche 1 bis 6, des Vektors nach einem der Ansprüche 9 bis 13, des Antikörpers oder des Fragments davon nach Anspruch 16 oder 17, der Antisense-RNA nach Anspruch 18 oder des Ribozyme nach Anspruch 19 zur Herstellung eines Arzneimittels zur Prävention oder Behandlung von Krankheiten, die mit einer Störung der Kontrolle der Genexpression in Zusammenhang stehen.

21. Verwendung des RNA-Moleküls nach einem der Ansprüche 1 bis 6, der DNA-Sequenz nach Anspruch 7 oder eines Fragments davon, des Antikörpers oder des Fragments davon nach Anspruch 16 oder 17, oder der Antisense-RNA nach Anspruch 18 oder eines Fragments davon zur Diagnose von Krankheiten, die mit einer Störung der Kontrolle der Genexpression in Zusammenhang stehen.
22. Verwendung nach Anspruch 20 oder 21, wobei es sich bei der Krankheit um eine Tumorerkrankung oder eine Erkrankung des Zentralnervensystems handelt.
23. Nicht-menschliches Säugetier, dessen NINTROX-Gen durch Insertion einer heterologen Sequenz verändert ist.
24. Nicht-menschliches Säugetier nach Anspruch 22, wobei die heterologe Sequenz eine Selektionsmarkersequenz ist.
25. Nicht-menschliches Säugetier nach Anspruch 22 oder 23, wobei die Selektionsmarkersequenz Resistenz gegen Neomycin vermittelt.
26. Verfahren zur Herstellung eines nicht-menschlichen Säugetiers nach einem der Ansprüche 22-25 gekennzeichnet durch die folgenden Schritte:
  - (a) Herstellung eines DNA-Fragments, insbesondere eines Vektors, enthaltend ein verändertes NINTROX-Gen, wobei das NINTROX-Gen durch Insertion einer heterologen Sequenz, insbesondere eines selektierbaren Markers, verändert worden ist;
  - (b) Präparation embryonaler Stammzellen aus einem nicht-menschlichen Säuger (bevorzugt Maus);
  - (c) Transformation der embryonalen Stammzellen von Schritt (b) mit

dem DNA-Fragment von Schritt (a), wobei das NINTROX-Gen in den embryonalen Stammzellen durch homologe Rekombination mit dem DNA-Fragment von (a) verändert wird,

- (d) Kultivieren der Zellen von Schritt (c),
- (e) Selektion der kultivierten Zellen von Schritt (d) auf das Vorhandensein der heterologen Sequenz, insbesondere des selektierbaren Markers,
- (f) Erzeugen chimerer nicht-menschlicher Säuger aus den Zellen von Schritt (e) durch Injektion dieser Zellen in Säuger-Blastocysten (bevorzugt Maus-Blastozyten), Übertragen der Blastozysten in pseudo-schwangere weibliche Säuger (bevorzugt Maus) und Analyse der erhaltenen Nachkommen auf eine Veränderung des NINTROX-Gens.

### **Zusammenfassung**

Beschrieben werden modular aufgebaute RNA-Moleküle, die an einen Liganden binden können und durch zwei Sequenzbereiche gekennzeichnet sind, einen ersten Sequenzbereich, der zur Aufrechterhaltung der dreidimensionalen Struktur des RNA-Moleküls beiträgt, und einen zweiten Sequenzbereich, der für die spezifische Bindung des Liganden verantwortlich ist. Diese RNA-Moleküle, beispielsweise die NINTROX-RNA, können zur direkten Beeinflussung der Genexpression verwendet werden. Beschrieben werden ferner die erfindungsgemäßen RNA-Moleküle enthaltende Vektoren sowie Arzneimittel und diagnostische Zusammensetzungen, die die vorstehenden RNA-Moleküle bzw. Vektoren enthalten, ein diese RNA-Moleküle spezifisch erkennender Antikörper bzw. spezifisch an diese RNA-Moleküle bindende Antisense-RNA oder diese RNA-Moleküle spaltende Ribozyme. Außerdem betrifft die Erfindung nicht-menschliche Säuger, deren NINTROX-Gen durch Insertion einer heterologen Sequenz verändert ist und daraus erhaltene Zellen.

Humane Sequenz des Nicht kodierenden RNA-Gens (einschließlich des putativen Promotors)

1 CTTAGAGTTT CGTGGCTTCA CGGTGGGAGT AGTTGGAGCA TTGGGGATGT  
51 TTTCTTACCA GACAAGCACA GTCAGGTTGA AGACCTAACCC AGGGCCAGAA  
101 GTAGCTTGC ACTTTCTAA ACTAGGCTCC TTCAACAAGG CTTGCTGCAG  
151 ATACTACTGA CCAGACAAGC TGTTGACCAG GCACCTCCCC TCCCCGCCAA  
201 ACCTTCCCCC CATGTGGTCG TTAGAGACAG AGCGACAGAG CAGTTGAGAG  
251 GACACTCCCCG TTTTCGGTGC CATCAGTGCC CCGTCTACAG CTCCCCCAGC  
301 TCCCCCCCACC TCCCCCAGTC CCAACCACGT TGGGACAGGG AGGTGTGAGG  
351 CAGGAGAGAC AGTTGGATTTC TTTAGAGAAG ATGGATATGA CCAGTGGCTA  
401 TGGCCTGTGC GATCCCACCC GTGGTGGCTC AAGTCTGCC CCACACCAGC  
451 CCCAATCCAA AACTGGCAAG GACGCTTCAC AGGACAGGAA AGTGGCACCT  
501 GTCTGCTCCA GCTCTGGCAT GGCTAGGAGG GGGGAGTCCC TTGAACACT  
551 GGGTAGAC TGGCCTGAAC CACAGGAGAG GATGGCCAG GGTGAGGTGG  
601 CATGGTCCCAT TCTCTGGGA CGTCCTCCAA CGGGTGGCGC TGGAGGCCAT  
651 GGAGGCAGTA GGACAGGTG CAGGCAGGCT GGCCTGGGGT CAGGCCGGGC  
701 AGAGCACAGC GGGGTGAGAG GGATTCCTAA TCACTCAGAG CAGTCTGTGA  
751 CTTAGTGGAC AGGGGAGGGG GCAAAGGGG AGGAGAAGAA AATGTTCTTC  
801 CAGTTACTTT CCAATTCTCC TTTAGGGACA GCTTAGAATT ATTTGCACTA  
851 TTGAGTCTTC ATGTTCCCAC TTCAAAACAA ACAGATGCTC TGAGAGCAAA  
901 CTGGCTTGAA TTGGTGCACAT TTAGTCCCTC AAGCCACCAG ATGTGACAGT  
951 GTTGAGAACT ACCTGGATTTC GTATATATAC CTGCGCTTGT TTTAAAGTGG  
1001 GCTCAGCACA TAGGGTTCCC ACGAAGCTCC GAAACTCTAA GTGTTGCTG  
1051 CAATTTTATA AGGACTTCCT GATTGGTTTC TCTTCTCCCC TTCCATTTC  
1101 GCCTTTGTT CATTTCATCC TTTCACTTCT TTCCCTTCCT CCGTCCTCCT  
1151 CCTTCCTAGT TCATCCCTTC TCTTCCAGGC AGCCGCGGTG CCCAACACAA  
1201 CTTGTGGCT CCAGTCCCCA GAACTCTGCC TGCCCTTGT CCTCCTGCTG  
1251 CCAGTACCAAG CCCCACCCCTG TTTTGAGCCC TGAGGAGGCC TTGGGCTCTG  
1301 CTGAGTCCAA CCTGGCCTGT CTGTGAAGAG CAAGAGAGCA GCAAGGTCTT  
1351 GCTCTCCTAG GTAGCCCCCT CTTCCCTGGT AAGAAAAAGC AAAAGGCATT  
1401 TCCCACCCCTG AACAAACGAGC CTTTTCACCC TTCTACTCTA GAGAAGTGG  
1451 CTGGAGGAGC TGGGCCCCGAT TTGGTAGTTG AGGAAAGCAC AGAGGCCTCC  
1501 TGTGGCCTGC CAGTCATCGA GTGGCCAAAC AGGGGCTCCA TGCCAGCCGA  
1551 CCTTGACCTC ACTCAGAAGT CCAGAGTCTA GCGTAGTGCA GCAGGGCAGT  
1601 AGCGGTACCA ATGCAGAAGT CCCAAGACCC GAGCTGGAC CAGTACCTGG  
1651 GTCCCCAGCC CTTCCCTCTGC TCCCCCTTTT CCCTCGGAGT TCTTCTTGAA

1701 TGGCAATGTT TTGCTTTGC TCGATGCAGA CAGGGGGCCA GAACACCACA  
1751 CATTTCACTG TCTGTCCTGGT CCATAGCTGT GGTGTAGGGGG CTTAGAGGCA  
1801 TGGGCTTGCT GTGGGTTTT AATTGATCAG TTTTCATGTG GGATCCCATC  
1851 TTTTTAACCT CTGTTCAGGA AGTCCTTATC TAGCTGCATA TCTTCATCAT  
1901 ATTGGTATAT CCTTTCTGT GTTACAGAG ATGTCTCTTA TATCTAAATC  
1951 TGTCCAACCT AGAAAGTACCT TATCAAAGTA GCAAATGAGA CAGCAGTCTT  
2001 ATGCTTCCAG AAACACCCAC AGGCATGTCC CATGTGAGGT GCTGCCATGA  
2051 ACTGTCAAGT GTGTGTTGTC TTGTGTATTT CAGTTATTGT CCCTGGCTTC  
2101 CTTACTATGG TGTAATCATG AAGGAGTGAA ACATCATAGA AACTGTCTAG  
2151 CACTTCCTTG CCAGTCTTTA GTGATCAGGA ACCATAGTTG ACAGTTCCAA  
2201 TCAGTAGCTT AAGAIIAAAC CGTGTGTTGTC TCTTCTGGAA TGTTAGAAC  
2251 TGAGGGAGTT TGCCCCGTTTC TGTTGTAGA GTCTCATAGT TGGACTTTCT  
2301 AGCATATATG TGTCCATTTC CTTATGCTGT AAAAGCAAGT CCTGCAACCA  
2351 AACTCCCATC AGCCCAATCC CTGATCCCTG ATCCCTTCCA CCTGCTCTGC  
2401 TGATGACCCC CCCAGCTTCA CTTCTGACTC TTCCCCAGGA AGGGAAGGGG  
2451 GGTCAAGAAGA GAGGGTGAGT CCTCCAGAAC TCTTCCTCCA AGGACAGAAG  
2501 GCTCCTGCC CCATAGTGGC CTCGAACCTCC TGGCACTACC AAAGGACACT  
2551 TATCCACGAG AGCGCAGCAT CCGACCAGGT TGTCACTGAG AAGATGTTA  
2601 TTTTGGTCAG TTGGGTTTTT ATGTATTATA CTTAGTCAAA TGTAATGTGG  
2651 CTTCTGGAAT CATTGTCCAG AGCTGCTTCC CCGTCACCTG GCGTCATCT  
2701 GGTCCCTGGTA AGAGGGAGTGC GTGGCCCACC AGGCCCCCCT GTCACCCATG  
2751 ACAGTCATT CAGGGCCGAT GGGGCAGTCG TGGTTGGGAA CACAGCATT  
2801 CAAGCGTCAC TTTATTTCAT TCGGGCCCCA CCTGCAGCTC CCTCAAAGAG  
2851 GCAGTTGCC AGCCTCTTTC CCTTCCAGTT TATTCCAGAG CTGCCAGTGG  
2901 GGCCTGAGGC TCCTTAGGGT TTTCTCTCTA TTTCCCCCTT TCTTCCTCAT  
2951 TCCCTCGTCT TTCCCCAAGG CATCACGAGT CAGTCGCCTT TCAGCAGGCA  
3001 GCCTTGGCGG TTTATCGCCC TGGCAGGCAG GGGCCCTGCA GCTCTCATGC  
3051 TGCCCCCTGCC TTGGGGTCAG GTTGACAGGA GGTTGGAGGG AAAGCCTTAA  
3101 GCTGCAGGAT TCTCACCAAGC TGTGTCCGGC CCAGTTTGG GGTCTGACCT  
3151 CAATTTCAAT TTTGTCTGTA CTTGAACATT ATGAAGATGG GGGCCTCTTT  
3201 CAGTGAATT GTGAACAGCA GAATTGACCG ACAGCTTCC AGTACCCATG  
3251 GGGCTAGGTC ATTAAGGCCA CATCCACAGT CTCCCCCACC CTTGTTCCAG  
3301 TTGTTAGTTA CTACCTCCTC TCCTGACAAT ACTGTATGTC GTCGAGCTCC  
3351 CCCCCAGGTCT ACCCCCTCCCG GCCCTGCCTG CTGGTGGGCT TGTCAAGGCC  
3401 AGTGGGATTG CCGGTCTTGA CAGCTCAGTG AGCTGGAGAT ACTTGGTCAC

3451 AGCCAGGCAG TAGCACAGCT CCCTTCTGTT GATGCTGTAT TCCCATATCA  
3501 AAAGGCACAG GGGACACCCA GAAACGCCAC ATCCCCAAT CCATCAGTGC  
3551 CAAACTAGCC AACGGCCCCA GCTTCTCAGC TCGCTGGATG GCGGAAGCTG  
3601 CTACTCGTGA GCGCCAGTGC GGGTGCAGAC AATCTTCTGT TGGGTGGCAT  
3651 CATTCCAGGC CCGAAGCATG AACAGTGCAC CTGGACAGG GAGCAGCCCC  
3701 AAATTGTCAC CTGCTTCTCT GCCCAGCTTT TCATTGCTGT GACAGTGATG  
3751 GCGAAAGAGG GTAATAACCA GACACAAACT GCCAAGTTGG GTGGAGAAAG  
3801 GAGTTTCTTT AGCTGACAGA ATCTCTGAAT TTTAAATCAC TTAGTAAGCG  
3851 GCTCAAGCCC AGGAGGGAGC AGAGGGATAC GAGCGGAGTC CCCTGCGCGG  
3901 GACCATCTGG AATTGGTTTA GCCCAAGTGG AGCCTGACAG CCAGAACTCT  
3951 GTGTCCCCCG TCTAACACACA GCTCCTTTTC CAGAGCATTC CAGTCAGGCT  
4001 CTCTGGCTG ACTGGGCCAG GGGAGGTTAC AGGTACCAAGT TCTTTAAGAA  
4051 GATCTTGGG CATATACATT TTTAGCCTGT GTCATTGCC CAAATGGATT  
4101 CCTGTTCAA GTTCACACCT GCAGATTCTA GGACCTGTGT CCTAGACTTC  
4151 AGGGAGTCAG CTGTTCTAG AGTCCTACC ATGGACTGGG TCTGGAGGAC  
4201 CTGCCCGCTG GGGGGGCAGA GCCCTGCTCC CTCCGGTCT TCCTACTCTT  
4251 CTCTCTGCTC TGACGGGATT TGTTGATTCT CTCCATTGGT GTGTCTTCT  
4301 CTTTAGATA TTGTATCAAAT CTTAGAAAA GGCAAGTCT ACTTGTATA  
4351 AATCGTTAGG ATACTGCCTC CCCCAGGGTC TAAATTACA TATTAGAGGG  
4401 GAAAAGCTGA ACACTGAAGT CAGTCTCAA CAATTTAGAA GGAAAACCTA  
4451 GAAAACATTT GGCAGAAAAT TACATTCGA TGTTTTGAA TGAATACAG  
4501 CAAGCTTTA CAACAGTGCT GATCTAAAAA TACTTAGCAC TTGGCCTGAG  
4551 ATGCCTGGTG AGCATTACAG GCAAGGGAA TCTGGAGGTA GCCGACCTGA  
4601 GGACATGGCT TCTGAACCTG TCTTTGGGA GTGGTATGGA AGGTGGAGCG  
4651 TTCACCAGTG ACCTGGAGG CCCAGCACCA CCCTCCTCC CACTCTTCTC  
4701 ATCTTGACAG AGCCTGCCCA AGCGCTGACG TGTCAGGAAA ACACCCAGGG  
4751 AACTAGGAAG GCACTTCTGC CTGAGGGGCA GCCTGCCTTG CCCACTCCTG  
4801 CTCTGCTCGC CTCGGATCAG CTGAGCCTTC TGAGCTGCC TCTCACTGCC  
4851 TCCCCAAGGC CCCCTGCCTG CCCTGTCAGG AGGCAGAAGG AAGCAGGTGT  
4901 GAGGGCAGTG CAAGGAGGGA GCACAACCCC CAGCTCCGC TCGGGCTCC  
4951 GACTTGTGCA CAGGCAGAGC CCAGACCCCTG GAGGAAATCC TACCTTGAA  
5001 TTCAAGAACCA TTTGGGAAT TTGGAAATCT CTTTGCCCCC AAACCCCCAT  
5051 TCTGTCCTAC CTTTAATCAG GTCCTGCTCA GCAGTGAGAG CAGATGAGGT  
5101 GAAAAGGCCA AGAGGTTGG CTCCTGCCA CTGATAGCCC CTCTCCCCGC  
5151 AGTGTGTTGTG TGTCAAGTGG CAAAGCTGTT CTTCCCTGGTG ACCCTGATTA  
5201 TATCCAGTAA CACATAGACT GTGCGCATAG GCCTGCTTTG TCTCCTCTAT

5251 CCTGGGCTTT TGTTTGCTT TTTAGTTTG CTTTTAGTTT TTCTGTCCCT  
5301 TTTATTTAAC GCACCGACTA GACACACAAA GCAGTTGAAT TTTTATATAT  
5351 ATATCTGTAT ATTGCACAAT TATAAACTCA TTTTGCTTGT GGCTCCACAC  
5401 ACACAAAAAA AGACCTGTTA AAATTATACC TGTTGCTTAA TTACAATATT  
5451 TCTGATAACC ATAGCATAGG ACAAGGGAAA ATAAAAAAAG AAAAAAAAGA  
5501 AAAAAAAACG ACAAAATCTGT CTGCTGGTCA CTTCTCTGT CCAAGCAGAT  
5551 TCGTGGTCTT TTCCCTCGCTT CTTCAAGGG CTTTCCTGTG CCAGGTGAAG  
5601 GAGGCTCCAG GCAGCACCCA GGTTTGACAC TCTTGTTCCT CCCGTGCTTG  
5651 TGAAAGAGGT CCCAAGGTTG TGGGTGCAGG AGCGCTCCCT TGACCTGCTG  
5701 AAGTCCGGAA CGTAGTCGGC ACAGCCTGGT CGCCTTCCAC CTCTGGGAGC  
5751 TGGAGTCCAC TGGGGTGGCC TGACTCCCCC AGTCCCCCTTC CCGTGACCTG  
5801 GTCAGGGTGA GCCCATGTGG AGTCAGCCTC GCAGGGCTCC CTGCCAGTAG  
5851 GGTCCGACTG TGTTCATCC TTCCCCTCT GTGAGGCTG GGGGCTGGAG  
5901 CGGAGACGGG AGGCCTGGCC TGTCTCGGAA CCTGTGAGCT GCACCAGGTA  
5951 GAACGCCAGG GACCCCAGAA TCATGTGCGT CAGTCCAGG GGTCCCCCTCC  
6001 AGGAGTAGTG AAGACTCCAG AARTGTCCCT TTCTTCTCCC CCATCCTACCG  
6051 ACTAATTGCA TTTGCTTTG TAATTCTTAA TGAGCAATAT CTGCTAGAGA  
6101 GTTTAGCTGT AACAGTTCTT TTTGATCATC TTTTTTTAAT AATTAGAAC  
6151 ACCAAAAAAA TCCAGAAACT TGTCTTCCA AAGCAGAGAG CATTATAATC  
6201 ACCAGGGCCA AAAGCTTCCC TCCCTGCTGT CATTGCTCT TCTGAGGCCT  
6251 GAATCCAAAA GAAAAACAGC CATAGGCCCT TTCAGTGGCC GGGCTACCCG  
6301 TGAGCCCTTC GGAGGACCAAG GGCTGGGGCA GCCTCTGGC CCACATCCGG  
6351 GGCCAGCTCC GGCGTGTGTT CAGTGTAGC AGTGGGTCA GATGCTCTT  
6401 CCCACCCAGC CTGGGATAGG GGCAGACGGAG GCGAGGAGGC CGTTGCCGCT  
6451 GATGTTGGC CGTGAACAGG TGGGTGTCTG CGTGCCTCCA CGTGCCTGTT  
6501 TTCTGACTGA CATGAAATCG ACGCCCGAGT TAGCCTCACC CGGTGACCTC  
6551 TAGCCCTGCC CGGATGGAGC GGGGCCACC CGGTTCACTG TTTCTGGGGA  
6601 GCTGGACAGT GGAGTGAAA AGGCTTGCAG AACCTGAAGC CTGCTCCTTC  
6651 CCTTGCTACC ACGGCCTCT TTCCGTTGA TTTGTCACTG CTTCAATCAA  
6701 TAACAGCCGC TCCAGAGTCA GTAGTCAATG AATATATGAC CAAATATCAC  
6751 CAGGACTGTT ACTCAATGTG TGCCGAGCCC TTGCCCATGC TGGGCTCCCG  
6801 TGTATCTGGA CACTGTAACG TGTGCCTGTG TTGCTCCCT TCCCTTCCCT  
6851 TCTTTGCCCT TTACTTGCTT TTCTGGGTT TTTCTGTTG GGTTGGTTT  
6901 GGTTTTTATT TCTCCTTTG TGTTCCAAAC ATGAGGTTCT CTCTACTGGT  
6951 CCTCTTAACG GTGGTGTGA GGCTTATATT TGTGTAATT TTGGTGGGTG

7001 AAAGGAATTT TGCTAAGTAA ATCTCTTCTG TGTGAACT GAAGTCTGTA  
7051 TTGTAACATAT GTTTAAAGTA ATTGTTCCAG AGACAAATAT TTCTAGACAC  
7101 TTTTCTTTA CAAACAAAAG CATTGGAGG GAGGGGGATG GTGACTGAGA  
7151 TGAGAGGGGA GAGCTGACA GATGACCCCT GCCCAGATCA GCCAGAAGCC  
7201 ACCCAAAGCA GTGGAGCCCA GGAGTCCCAC TCCAAGCCAG CAAGCCGAAT  
7251 AGCTGATGTG TTGCCACTTT CCAAGTCACT GCAAAACCAG GTTTGTTCC  
7301 GCCCAGTGGA TTCTTGTTT GCTCCCCTC CCCCCGAGAT TATTACCACC  
7351 ATCCCGTGCT TTTAAGGAAA GGCAAGATTG ATGTTCCCTT GAGGGGAGCC  
7401 AGGAGGGGAT GTGTGTGTG AGAGCTGAAAG AGCTGGGGAG AATGGGGCTG  
7451 GGCCCACCCA AGCAGGAGGC TGGGACGCTC TGCTGTGGC ACAGGTCAGG  
7501 CTAATGTTGG CAGATGCAGC TCTTCCTGGA CAGGCCAGGT GGTGGGCATT  
7551 CTCTCTCCAA GGTGTGCCCC GTGGGCATTA CTGTTAAGA CACTTCCGTC  
7601 ACATCCCACC CCATCCTCCA GGGCTCAACA CTGTGACATC TCTATTCCCC  
7651 ACCCTCCCCCT TCCCAGGGCA ATAAAATGAC CATGGAGGGG GCTTGCACTC  
7701 TCTTGGCTGT CACCCGATCG CCAGCAAAAC TTAGATGTGA GAAAACCCCT  
7751 TCCCATTCCA TGGCGAAAAC ATCTCCTTAG AAAAGCCATT ACCCTCATTA  
7801 GGCATGGTTT TGGGCTCCCA AAACACCTGA CAGCCCCCTCC CTCCTCTGAG  
7851 AGCCGGAGAG TGCTGACTGT AGTGACCATT GCATGCCGG TGCAAGCATCT  
7901 GGAAGAGCTA GGCAGGGTGT CTGCCCCCTC CTGAGTTGAA GTCATGCTCC  
7951 CCTGTGCCAG CCCAGAGGCC GAGAGCTATG GACAGCATTG CCAGTAACAC  
8001 AGGCCACCCCT GTGCAGAAGG GAGCTGGCTC CACCCCTGGAA ACCTGTCTGA  
8051 GGTTGGGAGA GGTGCACCTG GGGCACAGGG AGAGGCCGG ACACACTTAG  
8101 CTGGAGATGT CTCTAAAGC CCTGTATCGT ATTCAACCTTC AGTTTTGTC  
8151 TTTTGGGACA ATTACTTTAG AAAATAAGTA GGTCGTTTTA AAAACAAAAAA  
8201 TTATTGATTG CTTTTTTGTA GTGTCAGAA AAAAGGTCT TTGTGTATAG  
8251 CCAAATGACT GAAAGCACTG ATATATTTAA AAACAAAGG CAATTTATTA  
8301 AGGAAAATTTG TACCATTCA GTAAACCTGT CTGAATGTAC CTGTATAACGT  
8351 TTCAAAAACA CCCCCCCCCC ACTGAATCCC TGTAACCTAT TTATTATATA  
8401 AAGAGTTTGC CTTATAATT TA

Murine Sequenz des Nicht kodierenden RNA-Gens (einschließlich putativen Promotors)

1 CTTAGAGTTT CGTGGCTTCG GGGTGGGAGT AGTTGGAGCA TTGGGATGTT  
51 TTTCTTACCG ACAAGCACAG TCAGGTTGAA GACCTAACCA GGGCCAGAAG  
101 TAGCTTTGCA CTTTTCTAAA CTAGGCTCCT TCAACAAAGGC TTGCTGCAGA  
151 TACTACTGAC CAGACAAAGCT GTTGACCAGG CACTCCCCC AACAAATATCC  
201 TCCCTCTTCC CCCCCCCCCAC CCCCCGCCCCG TGTGCTCGTT AGGGCAATTG  
251 AAAGGACACT CCCATTTTG GTGCCATTGA TGCCCTGTCC ATAATAGCTT  
301 CCCTGACTTT TACACCACCC CAACTCCCCA TCTGAAGGAC TGGGAGGTGT  
351 GATGCCAGGAG AAACATATGGG ACTCTTGGGA GAAGACTATG GAGTTGGCCTA  
401 GTGATTAAGG CCCACTAATT CCAACTGTGG TAGCACAGAT CTGGCTCCAC  
451 ATCAACCCAA TCCAAAAC TG ACAAGGATAT TTTGCAAAAA AAGAAAAGTGG  
501 CACCTGTCTG ATCCAGCTCT GACATGGCTA GAGGTGAGTC CTAAACTGAT  
551 GGCTTATAAAA CTAGCCTGAG CCACAGAAGA GTATGGCCCA GAGTGAAGTG  
601 TCATCATCTG TTCACAAAGGC ATGCTCCCCT AGAAGATAAT GCTAAAGAGG  
651 TGCCATGGAG GCAGCAGGAC AAAGTACAGG CAGGCTAGGT GGAGTCAGC  
701 CAGGCCTAGT GCCACAGAAC AAGAGAGCAG TCTGACTAGT APTTAAGAGG  
751 GAAGAAAAGGA AAAATATTCTT CCAATTACTT TCCAGTTCTC CTTTAGGGAC  
801 AGCTTACAAT TATTTGCACT ATTGAGTCTT CATGTTCCCA CTTCAAAACA  
851 AACAGATGCT CTGAAAGCAA ACTGGCTTGA AATGGTGACA CTGTCCCACA  
901 AGCCACCAAGA CATGCCAGTG TTCAGAACTA CCTGTATCTG TATATACCTG  
951 CGCTTGTCTT AAAGTGGCT CAGCACATAG GATTCCCAAG AAGCTCCGAA  
1001 ACTCTAACGTG TTTGCTGCAA TTTTATAAGG ACTTCCTGAT TGCTTTCTCT  
1051 CTCGTCCTTC CATTCTTCC TTCCTTCCAT TTCATGCTTT CATTCTTCC  
1101 CCTAGCTTCT AGTTGTTCT TCTGTTCCAG GCAGCTGCAG TGCTGAACCA  
1151 CATGGTTACC TAACAGCAGT CAGCTGCAGC CCTAGGATTC TTCTGCCCT  
1201 TTAACTTCCC ATTGCCAGTG CCAGGTATCA TATTTAACCT TGAGCAAGAG  
1251 CTGGGCTCTT TTGAGCCCTC CCTAACCTCT GTGAAGAAGA ACAAGAAGGT  
1301 AGGAAGCTCT TGCTCTTGCT AAGAAAAATG TCAAAAGGCT TTCAGACCTT  
1351 AAACAATGAG CCTTTTCACC TTTTACTCTA GAAAAGTGGG CTAGAAAATC  
1401 TGGGTACAT TGGGTAGCTG AAGGAGATAC AGAGGCCCCCT ATGGCCTGCC  
1451 AGAGTCGTTG CATGGCCCAA CAGGGGCTCC ATGCCCACTA CCCTTGACCC  
1501 TACTCAGAAA TCTAATGTCA TACTTAGTGT GGGCAGGGGA CCTGTCAGGA  
1551 CAGATGCAGA CCTAACAGG GAGTGACACC AGGGCCCTTG GCCCTTCTTC  
1601 TGACAAACAT ACACATCCCA AGTCTTTTC TAGTGGAAATT CTTAACCTCT  
1651 TGCTCACTGG GGACTGGGAA GCATCAGCAC ATCCCCATATT TCAAAACTCTG

1701 CTCCATAAGT ACAGTGGTGA ATTTTATAGA CTTGACTTTG CTGTGGGTT  
1751 TTAATTGGTC AGTTTTAATT TGGGATCCCA AAGTTTTAAC CTCCATTCA  
1801 GAAGTCCTTA TCTAGCTGCA TATCTTCATC ATATTGGTAT ATCCTTTCT  
1851 GTGTTACAG AGATGTCTCA TATCTATCGA AATCTGTCTG AGAAGTACCT  
1901 TATCAAAGTA GCAAATGAGA CAGCAGTCTT ATGCTTCCAG AAACACCCAC  
1951 AGGCACGTCC CATGTGAGCT GCTGCCATGA ACTGTCGAGT GTGTATTGTC  
2001 TTGTGTATTT TCGTTAACGT TCCCCAGCTT CCTTCCTGCG GTGTAATCAT  
2051 GGAAGAGTGA AACATCATAG AAATCGTCTA GCACTTCCTG GCCAGTCCTT  
2101 AGTGATCAGG AACCGTAGTT GACAGTTCCA ATTGATAGCT TAAGATAAAA  
2151 CCATGTTTGT CTCTTATGGA ATGGTTAGAA CTAAGTGAGA GATCTTGCCC  
2201 CATTCTGTTT GCCGAATCAT AGTTGGACTT TTAGTGTATT TGTATCCATT  
2251 TCCTTGTGCT ATAAGAGCAA ACCCTGCAAC CAGCTTCTG TCAGGGAGTC  
2301 CTTTTGCCTG CTCTGCTTTT GATCCTCTTA GTCTTGCTTC TGGTTCCCTCC  
2351 CTGGAGAGGG AGGAGGGGTC AGAAGAGGAA TTCTGGAGGA TCCAGGATAT  
2401 GTCCTTCTGA ACTCCTGCTT CTTCCAGTGA CAAAGGGCCC CTACTGCC  
2451 ACCCCAAACCT GCCCCATGCA CTCCCTAGG ACACCTTCC ATACTTTCA  
2501 CAACACCTAG CCAGGTTGAC ACCAAGTTGT TTATTGTGGT CTGCTTGGAA  
2551 TTTTACCTGT TAGGCTTACT TAGTCCAATC AAATGGACTC CAAGTTGGGT  
2601 ATCCCTCATC TTTGGAGAC AACCTAGGCT GATTAGATAT TTACTTTGG  
2651 GATTGCAGCA CTTTGGGTGC CGTTTTCTT TTACTTGGGT TTTATCTGCA  
2701 GCTCCCTCAC CACCACCACC ACCCCCCACT TACCTGTATG TAGAACTGAT  
2751 TTCAAAACTG CAGGTGGTGG TAACTGCAGC TTCTTAGGGT TTTCTTC  
2801 TCTTGCTTCT TTCCCCATTC CCTCATCCAC AAATAAGGGC ATCACAAGTC  
2851 AGTCTCCTT AAGCAGGCAG CTTTGGTGGG GTTTTCCCC TGGAAGCCAG  
2901 GGACCCTGTC AGGCTGCCTC TGCTTGTGG TCAGGTTGAC AGGAGGTTGG  
2951 AGGGAAAAGC CTTAAGTCAT GGGATTCTCA CCAGCTGTGT CTGGCTCAGA  
3001 CCTGGAATGT GACCTTATT TTGTTGTATT TGAACATTGT AAAGTGTGGG  
3051 TGGTACCTTA AACTGAAAT GTGAAGAATC CAGAAACTGA CCAACAGCTT  
3101 TCAGATACCT GGGGCTAGGT CACTAAGGTC ACATCCAGTC TTCCCTACCC  
3151 TGTTCTACTT GTTAGCTACT ACCTCTCCCA GATAGATTGC TGTATATCCT  
3201 CCAACTATGA TCATCCTGGC CCAAGCTTGC CTGTTCTTGA GTCTGTCTTA  
3251 ACCAGTGGAA CTGCTGCCCT TGGTGTGCAG TGAGTTGAGG ACTCTTGGTC  
3301 ACAGCCAGGC TCTAGTAGTA CAGCTCCTTT CTGCTGGTGC TGTATTTCCA  
3351 TATCAAAGG CACAGGGGAG ATCTAGAAAT GCCATCTCCC CCAGTCCATC  
3401 AGTGCCAAAC AAGCCCATGA TCCCAGCATG GGTACAGACA ACTCTGTTCA

3451 GTGCTATCAC AACAGACTAG AGGCCATGAA CATTGGACGT GGGAAACCAGA  
3501 GCAACCCGAA TTGCTGCTGC TTTATTCCAGC TTTCCGTTGC TCTGACAATG  
3551 ATAAAACAAG GCAGTAACCTT AAAACAGACT GCCAGGTTTG GCAGAGAAAG  
3601 GAAATTCCCTT AGCTGACAGC ACCTCTGGAT TTTAAATAGG TTGTAATAAG  
3651 TGGCTCAAAC CCATCCAGGA AAAAGCAAAA GGGTTAGAAC TGACCAGATG  
3701 AGACCAGCCT GATTCATGC AGCCCAAATG GAGTCCAGCT GTCTGAACTC  
3751 TGCAGCACTT CTCTACTACA GTCTCCTAGA GCATTCCAGC CAGGCTCTTC  
3801 AGGCTGAGGA GACATCACAG GTGCCAGTTC TTCAAGAAGA CTTTTGTGCA  
3851 TCAGTTCATATA GCCTATATCT TTGCCAAGA TTGTAGATTC AGGTTAACAC  
3901 TACAGATTCT AGGGCAGATG ACTGAGACTC AGAAAAAAAG CCCCTGTGGA  
3951 CTGTTGGTATA GCGAAGTACA AAAACTGAAG GGGGCTAGGG CAGATGCCGC  
4001 ATGCCTCATG CCAGAGCCAA GCCCTCTGCT CCATCCACAT CCTTTCTGG  
4051 CTCCTTCTTC CTGCTCTCTG CTTCAAGTGAAC CCAGCCCCAC TCTGAAGAGA  
4101 TTTGTTGATT CTCTCCATTT TTATGTCTTT CTCTTTAGG TACTATATAG  
4151 AAAAGGCTTA GTCTAATTGT TATAAATTGC TAGAATACTG CCTCCCCCAG  
4201 GGTCTAAAAA TATATGCTAA AGGGGAAAAAC TTGAACACTG AAACCAGTTC  
4251 TGAACAATTT AGAAGGAAAAA CCTTGAAAAAC ATTTAACAAA AAATTATATT  
4301 TTAAATGTTA TGAATTAAGAG GAGGCTTTTG AAAAATGTT GATCTATAAA  
4351 TACTTACTTT AGGCCTGAGG TGTCTAATGA GTGAACGTGAG CAATGGGAAC  
4401 TCAAGGCTGA AGCCTCTGC ATCAGAGGAG GTAGAACCCAG GAGCCTCTTG  
4451 AGATTTGAGG TGTTTTAGCA TTGGAAAGCC ACTCTTGAG TAGCTGGCCC  
4501 CAGAAACTAC TTCTGACCTT GTCATTTGGA ATGGAGGTTA GTGGTCTGCC  
4551 AGATGCCAAA GCTGCATGAG ACCAGCTCTT GGTTTATCAA TTTGAACACT  
4601 CAGTAACCTA GAAGGCCAG CACAAAGTCT CTGCTCTCTT CTTAACTGAG  
4651 CCTGCCAG CACTACTGCA CAAATTAGGG AGGGTCTACT TCCTACAGAG  
4701 CATCCCTCCC TGGGCCCCCT CCCATCCTTT GTACTCTACC TACCTGACCT  
4751 TCAGGATCTT GGCACATACG AAATGGCTGT GTAGCAAGCA CTTTGGCATG  
4801 CCCTCCTAAA CTTACCCAG AGCCTCTCCC TGCCTCCTTA AGCCAGTCTG  
4851 CCTGTCTTCT GGGGAGGTGT TAGAGCCCAT AGAATGGAGA GGAGAAAGAA  
4901 AAGAGGAAGA GGCAGGCAGG TAGAAAAAG GCTCTGGAG GAAAGACAGC  
4951 CTCCTAGGCT TTGCACAAAGC AGGACTCAGC CCCTTGTGGG AACTAAGTGC  
5001 CATCTTGGAG TTTAAGAACAA TTTGGACAAG TTGCAAATGA CCTTTGCTCC  
5051 TTGCTCCTCT CACCTTTAT GGGGCCCTGC TTAGCACTGA AAGCAAATGC  
5101 GCTGAAAAGG CAAAGAGGTT TGGCTCCTGC CCACTGATAG TCCTTTCCCT  
5151 GCAGTGTGG TGTGTCAAGT GGCAAAGCTG TTCTTCCTGG TGACTCTGAT  
5201 TAGATCCAGT AACTTAAAGAG ATTGTATGC ATAGGTCTGC TTTGACTCTT

5251 CTATTCTGGG CTTTGATTT GTTTTCAGT TTTGCTTTA GTTTCCATAT  
5301 TTTTATTTA TGCACCAACT AGACACACAA AGCAGTTGAA TTTATATATA  
5351 TATATATATA TATATATCTG TATATTCAC AATTATAAAC TCATTTGCT  
5401 TGTGACGCCA CACACACACA AAAAGAAAAA CCTTTAAAAA TTATACCTGT  
5451 TGCTTAATTA CAATATTTCT GATAACCATA GAGTAGGACA AGGGAAAAAA  
5501 TTTAAAAAAA AAAAAAAA AAGAAAAAAC ACATCTGTCT GCTGGTCACT  
5551 TCTTCAATCC AAGCAGATCT GTGATCTTTC CTCGCGTCTT TCAAAGACTT  
5601 CCCTGTGCTA AGTGAAGGAA GCTCCAGGCT GCACCCAGGT TTTGTGCTTT  
5651 GTTTCTCCTC TGTTGTGAAA GGGGCCCCAA GATTCTGGGT ACAGGACAGT  
5701 TCATTTCAGC ATGGGGTCAG GAGACAAGAG CACTCCCTTT ACATGCTGAC  
5751 GTACAGAACT TAGTCGGAAT AGCCTAGTCC CCACCTCTAG GGATGGGGAG  
5801 CTAGGCATGCA TGGGGGTGAC CCAACTCCCT CCACCTTCC CTGGCCAGGA  
5851 AGAGCCTGTG TACAGTAAGT CTGACAAGCT TTCCCCAGTT AGCAGGGCTC  
5901 AGAGCATTAA AAAACCTCC AAACTTGCT GAGTCTAGGG ACTAGAGAGA  
5951 AGATAGAAGA TTTGGTCTAT CTCCAAGGTG TGTAAGCTGT ACCAGGTAGA  
6001 ATGCCAGGGAA CCCCAGAACC ACATCCAACA GCCCAATGGG TCTCCTCCAG  
6051 AAAGTAGTGA AGACTCCAGA AACATCCCTT TCTCTCTCC CTGCTCCCAT  
6101 GAGTAACTGC ATTTGCTTTT GTAATCCTTA ATGAGCATTAA TCTGCTAAAA  
6151 AAAAAAAATT AGCTGTAACA GTTCTTTTG CAAAAGGATC ATTCTTAAAT  
6201 AATTAAAAAC ACCCCCCCCC CAAAAAAAAG TCCAGAACCT TGTTCTTCCA  
6251 AAGCAGAGAG CATTATAATC AGGGCCAAAA TCTGTCCCAC ACCTCTACCC  
6301 CATCTCCTCA TGATTGCTGC TTCTAAGGCC AGAATACAGC AAAGATATTT  
6351 GTAGGCCCTT TGGGTGACTG GGCTACCCCTT GGAGCTCTTG GAAGATGGC  
6401 TGGGAAAGCC TCTGAGACCC TATCCTAGGG CCTTGCTCTA GGGAGTAATC  
6451 AGTATTAGTA GAGTGTACCA ACATTATTCC CCAGCCGGCA TGAGATGGGG  
6501 GCAGAAGAAG CCAAAGGTT GTCTCCACTG CTACTTACTT GGCCACTGAC  
6551 AGGTAGGTGA CCATGTATGT CCATATGCAT GTTTTATGGC TGATGTGAGA  
6601 TCAGCACCCA AGTTAGCTTC ACCTGGTGAC CTCTAACCCCT GCCTGGATGG  
6651 AGCAGGCCAC CTGGTTCAAT GTTTCTGGGC AGCTGGACAA TGGAGTGCAA  
6701 AAGGCTTACA GAACTGAAG CCTTTCCCTT ACTTTGCTAG CACGGCCTCC  
6751 TTTTCCATTG GATTGTCAC TGCTTCAGTC AATAACAGCC GCTCCAGAGT  
6801 CAGTAGTTGA TGAATATATG ACCAAATATC ACCAGGACTG TTACTCAACG  
6851 TGTGCCGAGC CCTTTCCCTTG TGCTGGCTC CCTGTGTACC TGGACACTGT  
6901 AATGTGTGCT GTGTTGCTC TCCTTCCTCT TCCTTCCTTG CCCTTCCCTT  
6951 GTCTTCTGG GGTTTTCTG TTGGGTTTGG TTTGGTTTA TTTTCCTT

7001 TGTGTTCCAA ACATGAGGTT TTCTCTACTG GTCCTCTTA ACTGTGGTGT  
7051 TGAGGCTTCT ATTTGTGAA TTTTGTTGG GTGAAAGGAA CTTTGCTAAG  
7101 TAAATCTCTT CTGTGTTGA AATGAAGTCT GTATTGTAAC TATGTTAAA  
7151 GTAATTGTTTC CAGAGACAAA TGCTTCTAGG TACATTTCA TTACAAACAA  
7201 AGCATTGAA GGGAGGGAAAG TGGTGAATAA GACAAGAGGG GCAATCTGAA  
7251 TTGATCCCTG CCCAGATCAG CCAGAAGCTA CCAAAAGTTA AGCACTGGTT  
7301 TTCCATTCCA AGTCAAGAGA CTGAAGCTGA TGTTTGCCA TTTCAAAGT  
7351 CAAAGCAAAA CCAGCTTTC CACCAATGG ATTCTTGCT TCTCCTTCCC  
7401 AGATTATTAC TACTGCTGTA ATAATCTAGG AGTGCCAGGA GGGAAAGGAG  
7451 TATTAACACA GAGCTGTGCT CACTGAGTAT GGAAAGGCTT GGTCTGAGTT  
7501 TTCAGGAGGA TGACCCACTG TGGACATGGG GAGAAGACAG AAGATAAAATT  
7551 AGCCGCTCCC TGCCTAAGAT ACCTCTTAAAT AGATAAGTCA AGGCCATGGA  
7601 CATTATTGTC TACAAGGCAT GTTCAAAGA CATGACCAGT CAGGACACTT  
7651 CTGTCATACT CCATGTTGCC CCCTAGTACA CAGTACTAAT CTGATATCTC  
7701 TGTTCCCGCC ATGCCTGGGG GATAAAATGA TAGCAGAGAC TCCCTTCCTT  
7751 CAATGTGATC TAATTCCCAA CAAATCTGG GCCTGAGATA CCACCTGTTT  
7801 CTATGGCAAA CATCCTCAGT AAAGTGTAT TCTCATTGCA GATTGTTCCA  
7851 GCCTAATGTA AGAGGAAACAG AGCACTGTTG CCTTGGAGCC TCATGTGGAC  
7901 AGTTCTACCT GTAGTGACCA GTTGGCTATA GTAGTTATTA GCTGGAACAA  
7951 CCAGACAGGG TACATCCCC CTCCAAAATC CATGTTGTAC TCCCCTCTGC  
8001 CAGCCAGGGGG GGGTGAGATC TGTAGAATAG TGCAGCCAGT GACAAGCCAC  
8051 CTTGTGTTTG TCACCAGCTC AAAAATCAT CTAAGGTTGG GAGCAGGCAG  
8101 ACAAGGCAGA GAGAAAGATC CAGGACAGAC CTAGCTGGC TGGAGGGTC  
8151 TTGAAAAGCC CTCTGTCGTA TTCACCTTCA GTTTTGTC TTTGGGACAA  
8201 TTACTTTAGA AAATAAGTAG GTCGTTTAA AAACAAAATA TTGATTGCTT  
8251 TTTTGTAGTG TTCAAAACAA AAGCTTCTTT GTGTATAGCC AAATGACTGA  
8301 AAGCACTGAT ATATTTAAA ACAAAAGGCA ATTTATTAAG GAAATTTGTA  
8351 CCATTTCACT AAACCTGTCT GAATGTACCT GTATACGTTT CAAAAACACA  
8401 CCCCCACTGAA CCCCTGTAAC CTATTTATTA TATAAAGAGT TTGCCTTATA  
8451 AATTTACATA AAAA

1849 TCTTTTAACTTTCAGGAAGTCCTTATCTAGCTGCATATCTTCATC  
1781 AAG-----CA  
1899 ATATGGTATATCCTTTCTGTGTTACAGAGATGTCCTTAA..TATCTA  
1831 -----A-TC-----G-  
1947 AATCTGTCCAACGTAGAGAATACCTTATCAAAGTAGCAAATGAGACAGCAG  
1881 -----  
1997 TCTTATGCTTCCAGAAAACCCACAGGCATGTCCCATGTGAGCTGCTGCC  
1927 -----C-  
2047 ATGAACCTGTCAGTGTGTTGTCCTGTGTTAGTATTG..TCCCTG  
1977 -----G-----A-----TC-----AC-T-----CA  
2096 GCTTCCTTACTATGGTGTAAATCATGAAGGGTGAACACATCATAGAAAAGT  
2027 -----C-GC-----G-A-----TC-  
2146 TCTAGCACTTCCCTGCCAGTCCTTAGTGTAGGAAACCATAGTTGACAGT  
2077 -----G-----C-----G-  
2196 TCCAATCAGTAGCTTAAGAAAAACCGTGTGTTCTCTCTGGAAATGGTT  
2127 -----TGA-----A-----A-----A-  
2246 AG...AAAGTGAGGGAGTTGCCCCGTTCTGTTGTAGAGTCTCATAGTT  
2177 --AACT-----A-TC-----A-----CC-----A-  
2292 GGACTTCTAGCATATGTGTCATTTCCTTATGCTGTAAGAACAGTC  
2225 -----TG-----T-----A-----G-----A-----AC-  
2342 CTGCAACCAAACCTCCCATCAGCCCAATCCCTGATCCCTGATCCCTTCCAC  
2274 -----GCT-T-TG-----G-----.....GT-----TTG-  
2392 CTGCTCTGCTGATGACCCCCCAGCTTCACTTCTGACTCTCCCCCAGGAA  
2308 -----TT-----T-TT-----TC-TG-----GT-----C-----TG-AG-  
2442 GGGAGGGGGTCAAGAGAG.....AGGGTGAGTCCCTC  
2358 -----G-A-----.....GAATTCTGGAGGATCC-----A-AT-----T-  
2476 AGAACT...CTTCCTCAAGGACAGAACAGGCTCTGCCCCCATAGTGGCC  
2408 T-----CCTG-----T-----GT-----A-----C-----A-TG-----CC-  
2522 TCGAACT...CCTGGCACTACCAAAGGACACTTATCCA.CGAGAGCGCAG  
2454 C-A-C-GCC-----AT-----C-TCT-----C-T-----TACCTTT-A-A  
2568 CATGCCACCGAGGTTCTCACTGAGAAGATGTTTATTTGGTCAG..TTGGGT  
2504 --C-TAG-----A-----C-----T-----G-----T-C-----AA  
2617 TTTTATGTATTA...TACTTAGTCAAATGTAATGTCGGCTCTGGAAATCA  
2551 -----C-G-----GGCT-----C-----CA-  
2663 TTGTCACAGAGCTGCTTCCCCGTACCTGGGCGTCATCTGGTCTGGTAAAG  
2586 -----AC-----AA-----TGGG-----ATCCC-----.....T-G-  
2713 AGGAGTGGCGTGGCCCCACAGGGCCCCCTGTCAACCATGACAGTTCATTCA  
2619 -----.....AC-T-  
2763 GGGCCGATGGGGCAGTCGTTGGGAACACAGCATTCAAGCGTC..ACT  
2626 A-----T-----TA-AT-T-TACTT-----TTG-----C-----TGG-T-C-GTT-  
2812 TTATTCTATCGGGCCCCACCTGCACTGCTCCCTCAAAGAGGGCAGTGGCCA  
2676 -----C-----T-C-T-----TTT-T-----..CCAC-----CCAC-A-C  
2862 GCCTCTTCCCT...TCCAGTTATTCCAGAGCTGCCAGTGGG...C  
2724 C-CAC-----A-----GTATG-AG-AC-G-----T-----A-A-----AG-----T-GTAA  
2904 CTGAGCTCTTACGGGTTCTCTCTATTTCCTTCTTCTCATTCC  
2774 -----CA-----T-----.....TC-C-----TTG-----CT-----C-  
2954 CTCGCTTCCCCAA...GGCATCACAGACTGTCGGCTTCAAGCAGG  
2822 -----A-----TAAG-----A-----T-----A-  
3000 AGCCTTGG..CGGTTTATGCCCTGGCAGGCAGGGGCCCTGCAGCTCTCAT  
2869 -----T-----TG-----G-----T-TC-----A-----C-----A-----G  
3049 GCTGCCCTGCTTGGGTCAAGGGTACAGGGAGGTTGGAGGG..AAAGCCT  
2913 -----T-----T-----.....A-  
3098 TAAGCTGCAGGATTCTCACCAAGCTGTGTCGGGCCAGTTTGGGCTCTGA  
2963 -----TCATG-----T-----T-----ACC-----AA-G-  
3148 CCTCAATTCAATTGTCTGTACTTGAACATTATGAA..GATGGGGCC  
3013 -----TT-----T-----.....G.A..GT..TG..T..TA  
3196 TCTTCAGTGAATTGTGAACAC..GCAG..AATTGACCGACAGCTTCCAG  
3056 C-----AA-C-----A-----G-ATC-----A-C-----A-----AGA  
3243 TACCCATGGGCTAGGTCAATTAGGCCACATCCACAGTCTCCCCCACCT  
3106 -----C-----T-----.....T-----A-C-----T-----C-  
3293 TGTTCACTGTGTTAGTACTACCTCTCTCTGACAAATACTGTATGTCTG  
3151 -----T-----C-----.....TC-----CAGAT-G-T-----G-----A-C-  
343 CGAGCTCCCCCAGGTCAACCCCTCCGGCCCTGCCTGCTGGTGGGCTTG  
201 -----C-A-----AT-A-AT-----TGG-----AAG-T-----T-CT-----A-TC-  
393 TCATAGCCAGTGGGATGGCCCTTGTGACAGCTCAGTGAGCTGGAGATAC  
246 -----T-----A-C-----T-CC-----GT-TG-----T-----AG-----CT-  
443 TTGGTCACAGGCCAGGCC..TAGCACAGCTCCCTCTGATGATGCTGTA  
295 -----T-----TAG-----T-----.....T-----C-G-  
490 TTCCCCATATCAAAGGCACAGGGACACCCAGAAACGCCACATCCCCAA  
345 -----T-----.....G-T-T-----T-----TC-----G-  
540 TCCATCAGTGCACAAACTAGCCAACGGCCCGAGCTCTCAGCTCGCTGGAT  
395 -----A-----C-T-AT-----A-----.....  
590 GCGGAAAGCTGCTACTCGTGAGCGCCAGTGCAGGGTGAGACAATCTTCCTG  
430 -----.....A-----C-----.....C-  
640 TTGGTGGCATATCCAGGCCGAAG..CATGAACAGTGCACACTGGGACAC  
447 -----CA-----CT-----CAA-----A-TA-----G-C-----T-----G-----AC

3689 GGGAGCAGCCCCAAATTGTCACCTGC . . . . . CTGCCAGCTTTTCAATTGCT . . . . .  
 3497 CA-----A-G-T-----G-----ATT-----C-G-----  
 3739 CTGACAGTGTGGCGAAAGGGTAAATGCCAGACACAAACTGCCAAGTT  
 3542 C-----A-----A-C-A-C-G-----TTA-A-----G-----  
 3789 GGGTGGAGAAGGAGTTCTTAGCTGACAGAATCTCTGAATTAAATC  
 3589 T-----CA-----AA-----C-----C-----G-----A-----  
 3839 ACT . . . TAGTAAGCGGCTCAAGC . . . . . CCAGGGAGGAGCAGAGGGATACGA  
 3639 GG-TG-----A-----T-----A-CCAT-----AAA-----A-A-G-TA-----  
 3883 GCGGAGTCCCCGTGGCGGGACCATCTGGATTGGTTAGCCAAAGTGGG  
 3689 A-T-----CAGAT-----A-----G-CT-----T-----CA-GC-----A-----  
 3933 CCTGACAGCCAGAACTCTGTGCCCCGCTAACACAGCTCTTCTTCCA  
 3734 T-CAG-T-T-----T-----CAG-A-TTC-----T-----C-C-T-----  
 3983 GAGCATTCAGCTCAGGCTCTCTGGGCTGACTGGGCCAGGGAGGTTACAG  
 3779 -----C-----TCA-----A-----CA-C-----  
 4033 GTACCAAGTCTTAAGAAGATTTGGGCATATACATTTAGCCTGTG  
 3821 -----G-----C-----CT-----T-----CAG-----CA-----A-A-----  
 4083 CATTGCCCAATGGATTCTCTGGTCAAGTCACACCTGCAGATTCTAGG  
 3869 T-----A-----GTA-A-----G-----A-----  
 4133 ACCTGTGCTCTAGACT . . . . . TCACGGAGTCAGCTGTTCTAG  
 3914 G-AGA-----A-TG-----CAGAAAAAAAGCC-----CT-TG-----A-T-TG-----A-AGC-----  
 4171 AGTCCTACCATGGAGTGGGCTGGAGGA . . . . . CCTGCCGGTGGGG  
 3964 -AG-A-A-AA-C-----AG-----G-A-G-C-GATGCCG-----A-----TCA-----CCA-----  
 4214 GGGCAGAGCC . . . CTGCTCCCTCC . . . . . GGGTCTTCTACTCT  
 4014 A-----CA-----CT-----A-----ACATCCTTTCT-----C-C-T-----T-----  
 4250 TCTCTCTG . . . . . CTCTGACGGGATTGTTGATTCT  
 4063 G-----CTTCAGTGAACCGAGCCCA-----A-----  
 4281 CTCCATTGGTGTCTTCTCTTTAGATATTGATCAATCTTAGAAAA  
 4113 -----TA-----G-----C-A-----  
 4111 GGCATAGTCTACTTGTATAATCGTTAGGATACTGCCCTCCCCAGGGTC  
 4056 -----T-----A-----T-C-----A-----  
 4205 -----A-----T-----GC-----A-----CT-----AC-----G-----  
 4431 CAATTAGAGGAAACCTAGAAAACATTGGCAGAAAATTACATTCGA  
 4255 -----T-----AA-----A-----T-----TAA-----  
 4481 TGTGTTGGAATGAAATACAAGCAAGCTTTACAACAGTGTGATCTAAAAA  
 4305 -----AG-----G-G-----GA-----A-----A-----T-----  
 4531 TACTTAGCACTGGCTGAGATGCCGTGGTAGCATTACAGGCAAGGGAA  
 4351 -----TT-----A-----G-T-----AA-----T-----TGAATCT-----T-----  
 4581 . . . . . TCTGGAGGTAGCCGACC  
 4450 GAGATTGAGGTGTTTACCATGGAAAGCC-----TTG-----T-G-----  
 4598 TGAGGACATGGCTCTGAACCTGCTTTGG . . . . . GAGTGGTATG  
 4500 CC-----A-CTA-----C-T-----A-----ATGGAGGT-----C-----  
 4639 GAAGGTG . . . . . GAGCG  
 4549 CC-----A-----C-----AAACAGTCATGAGACCAGCTTGGTTATCAATT-----A-A-----  
 4651 TTACCAAGTGACCTGGAAGGCCAGCACACCCCTCTTCCACTCTCTC  
 4599 C-----A-----A-----A-----AGTGT-----G-----  
 4701 ATCTTGACAGAGCTGCCAGCGCTGACGTGTCAAGGAAACACCCAGGG  
 4639 T-----A-----T-----A-----A-----T-----TGCACA-----  
 4751 AACTAGGAAGGCACTCTGCTGAGGGCAGCCCTGCCCT . . . . . GCCCCATCC  
 4673 -----T-----G-----CT-----A-----ACA-----A-----T-----CT-----C-GG-----C-----  
 4799 . . . . . TGCTCTGCTGCCCT . . . . . CGGA  
 4823 CATCCTTG-A-----A-CTA-----GACCTTCAGGATCTGGCACATA-----A-----  
 TCAGCTGAG . . . . . CTTCTGAGCT . . . . . GG  
 4815 ATG-----T-----TAGCAAGCACTTGGCATGC-----C-----A-----TACCCAGA-----  
 4823 CCTCTCACTGCCCTCCCCAAGGCCCTGCCCTGCCCT . . . . .  
 4823 -----C-----TT-----C-ACT-----T-----T-----CTGGGAGGTGTTA-----  
 4875 . . . . . GTCAAGGGAGCAGGAAGCAGGGTG  
 4873 GAGCCCATAAGGGAGGGAGAA-A-----A-----A-----G-C-----A-----  
 4900 TGAGGGCAGTGCAGGAGGGAGCACACACCCAGCTCCGCCCTGGGCTC  
 4923 GT-----AAAG-----CT-----TG-----A-----AG-----G-----T-----TAGG-----  
 4950 CGACTTGTGACAGGCCAGGCCAGGGCTGGAGG . . . . . AAATCTTAC  
 4960 -----T-----A-----GA-----T-----C-----T-----T-----GAAC-----G-G-----C-T-----  
 4995 TTGAAATTCAAGAACATTGGGAAATTGGAAATCTCTTGGCCCTTAAAC  
 5005 -----G-----G-----T-----AC-----G-----C-----GA-----C-----TG-----T-----TTG-----  
 5045 CCCCATCTGCCCTACCTTAACTCAGGCTCTGCCAGCTGAGAGCAGA  
 5055 T-----T-----TC-----T-----GG-----C-----T-----C-----A-----A-----  
 5095 TGAGGTGAAAGGCCAGAGGTTGGCTCTGCCACTGATAGCCCTCT  
 5098 -----C-----C-----A-----T-----A-----T-----T-----  
 5145 CCCCGAGTGTGTTGTGTCAGTGGCAAAGCTGTTCTCCCTGGTGACCC  
 5147 -----T-----T-----  
 5195 TGATTATATCCAGTAACACATAGA . . . . . CTGTGGCATAGGCCCTGCTTGT  
 5197 -----G-----TT-----GATT-----AT-----T-----A-----  
 5242 CTCCCTATCTGGCTTGTGTTGCTTGTGTTAGTTGCTTTAGTTT  
 5246 -----T-----T-----A-----T-----C-----  
 5292 TCTGCCCCCTTTATTAGCAGCAGGACTAGACACACAAAGCAGTTGAATT  
 5296 C-----A-----T-----T-----A-----  
 5342 . . . . . ATATATATATCTGTATATTGCAACATTATAAAACTC  
 5343 TATATATATA . . . . . T-----  
 5380 ATTTTGCTTGTGGCTCCACACACACAAAAAG . . . . . ACCTGTTAAATT  
 5393 -----A-----G-----C-----AAAA-----T-----  
 5426 ATACCTGTTGCTTAATTACAATATTCTGATAACCATAGCATAGGACAG  
 5443 -----AA-----CT-----A-----G-----  
 5476 GGAAAATA . . . . . AAAAAGAAAAAAAGAAAAAAAGCACAATCTGCTGC  
 5493 -----A-TT-----A-----A-----G-----AA-----C-----  
 5525 TGGTCACTTCTCTGTCCAAGCAGATTGCTGGCTTTCTCGCTTCTT  
 5543 -----AA-----CT-----A-----G-----  
 5575 CAAGGGCTTCTGTGCCAGGTGAAGGAGGCTCCAGGCAGCACCCAGGG  
 5592 -----A-----A-----C-----T-----A-----T-----  
 5625 TTGCACTCTGTTCTCCCGTCTGTGAAAGAGGTCCCAAGGTTCTGGG  
 5642 -----TG-----TC-----G-----C-----A-----  
 5675 TGCAG . . . . . GAGCGCTCCCT  
 5690 -----A-----GACAGTTCAATTCAAGCATGGGTCAAGGAGACAA-----A-----  
 5692 GACCTGCTGAAGTCGGAAAGCTGGTAGCAGCCCTGGCTTCCACCC  
 5740 T-----A-----C-----A-----T-----G-----A-----T-----A-----C-----  
 5742 TCT . . . . . GGGAGCTGGACTGCACTGGGTGGCTGACTCCCCAGTC  
 5786 -----AGGGATG-----A-----CA-G-----TG-----A-----CA-----T-----  
 5785 CCCCCCGGTGACCTGGTCAGGGTGAAGCCCATGTGGAGTCAGCCCTCGCAG  
 5833 A-----T-----C-----AA-----TG-----AC-----A-----T-----GA-----A-----  
 5835 GCCT . . . . . CCCCTGCAAGTAGGG . . . . . TCCGAGTGTGTTCATCTTCC . . . . . CACTCT  
 5878 -----T-----TC-----A-----TT-----C-----A-----CA-----T-----AAA-----A-----C-----AA-----T-----  
 5881 GTCGAGCCTGGGGCTGGAGCGAGACGGGGGGCTGGCTGCTCGG  
 5928 -----CT-----A-----A-----A-----A-----TA-----A-----ATT-----T-----A-----CA-----G-----  
 5930 ACCTGTGAGCTGCACCAAGGAGAACGCGAGGGACCCCCAGAATCATGTGG  
 5978 GTG-----A-----T-----T-----C-----CA-----C-----CA-C-----A-----  
 5980 TCAGTCAAGGGTCCCTCCAG . . . . . GAGTAGTGAAGACTCCAGAAATGTCC  
 6028 A-----C-----T-----T-----AA-----CA-----  
 6029 CTT . . . . . TCTTCTCCCCCATCTACAGTAATTGCAATTGCTTTGTAATT  
 6078 -----TC-----T-----T-----G-----C-----T-----C-----  
 6077 TTAATGAGCAATATCTGT . . . . . AGAGAGTTAGCTGTAACAGTTCTT  
 6128 -----T-----AAAAA-----A-----A-----AA-----  
 6122 TTG . . . . . ATCATCTTTTTAATAATTAGAAACACC . . . . . AAAA  
 6178 -----CAA-----GG-----A-----C-----A-----A-----CCCCCCCCAA-----  
 6158 AAATCCGAAACATTGTTCTCCAAGCAGAGAGCATTATAATCACCAGGG  
 6228 -----G-----C-----  
 6208 CCAAAAGCT . . . . . TCCCTCCCTGCT . . . . . GTCAATTGCTTCTTCT  
 6275 -----T-----G-----A-----A-----CT-----ACCCCATCTCCTCA-----G-----  
 6244 GAGGCCTGAATCCAAAAGAAAAACAGCCATAGGCCCTTCTAGTGGCCGG  
 6325 A-----A-----C-----GC-----G-----T-----TTG-----GG-----A-----T-----  
 6294 CTACCCGTGAGCCCTTCCGGAGGACCGGGCTGGGCAGGGCTGGGCCCA  
 6373 -----T-----G-----TC-----T-----A-----T-----A-----A-----A-----C-----  
 6344 CATCC . . . . . GGGGAGCTCCGGCTGTGTTCAAGTGTAGGGTCATG  
 6421 T-----TA-----C-----TT-----TA-----G-----A-----AA-----A-----T-----A-----T-----CA-----  
 6392 ATGCTCTTCCACCCAGCCTGGGATAGGGCAGAGGAGGGCAGGGAGGCC  
 6471 -----CAT-----A-----C-----G-----G-----A-----A-----A-----A-----C-----AAG-----TT  
 6442 GTTGGCGCTG . . . . . ATGTTGGCGTGAACAGGTGGGTTCTGCTGCTGGCGT  
 6521 -----CT-----A-----T-----ACT-----AC-----T-----A-----CAT-----AT-----  
 6488 CCACGTGCGTGTGTTCTGACTGACATGAAATCAGCAGCCCGAGTTAGCC  
 6571 -----TA-----A-----G-----TG-----G-----AG-----A-----A-----T-----  
 6538 ACCCGGTGACCTCTAGCCCTGCCGGATGGAGCGGGGCCACCCGGTTCA  
 6621 -----T-----A-----T-----A-----T-----A-----T-----  
 6588 GTGTTCTGGGAGCTGGACAGTGGAGTGCAAAAGGCTTGCAAGACTTGA  
 6669 A-----C-----A-----  
 6638 AGCCCTGCTCTCCCTGCTACACGGCCCTCC . . . . . TTTCGGTTGATTGTC  
 6719 -----TT-----A-----T-----G-----T-----A-----  
 6687 ACTGCTTCAATCAATAACAGCCGCTCCAGAGTCAGTGTCAATGAATATA  
 6769 -----G-----TG-----  
 6737 TGACCAAATACTACCAAGGACTGTTACTCAATGTGTGCCAGGCCCTTGCC  
 6819 -----C-----  
 6786 CATGCTGGGCTCCC . . . . . GTGTATCTGGACACTGTAACGTGTGCTGTTG  
 6869 TG-----T-----C-----  
 6835 TCCCCCTCCCTCTTCTGCTTCTGCTTCTTACTGTCTTCTGTTGGGTTT  
 6919 -----T-----C-----  
 6885 TGTGTTGGGTTGGTTGTTTATTCTCTTGTGTTGCTTCAACATG  
 6969 -----T-----  
 6935 GGTCTCTACTGGCTCT . . . . . TTAACTGTGGTGTGAGGCTTATATTG  
 7017 -----T-----  
 6984 GAAATTTGGGGTGGGAGAGGAATTGCTAAGTAAATCTCTCTG  
 7067 -----C-----  
 7034 TTGAACGTGACTCTGTTGTAACATGTTAAAGTAATTGTTCAAGAGA  
 7117 -----A-----  
 7084 CAAATTTCTAGACACTTTCTTACAAACAAAGCATTGGAGGGAG  
 7167 -----GC-----GT-----A-----A-----T-----A-----

7134 GGGGATGGTGA~~TGAGATGAGAGGGG~~CTGAACAGATGACCCCTGCC  
 7216 --AAG---A-A---CA---CA-T---T---T---  
 7184 CAGATCAGGCCAGAAGGCC~~AAAGCAGTGGAGCC~~CAGGAGTCCC~~ACTCC~~  
 7263 ---T---A---TA---A-T---TT-T---T---  
 7234 AAGCCAGCAAGCCGA~~ATAGCTGATGTTGCC~~ACTTCC~~AACTGCA~~TGCA  
 7310 ---T---AG-GA-T---T---T---A---AA---  
 7284 AAAC~~AGGTTTGTCCGCC~~AGTGGATTCTGTTGCT~~CCCC~~CTCCCC  
 7358 ---C---A---A---T---T---  
 7334 CCGAGATT~~ATTAC~~ACCATCCCGTGTCTTTAAGGAAAGGCAAGAT~~TGATG~~  
 7401 ...-T---?---G-A---  
 7384 TTTCTTGAGGGGAGCC~~CAGGAGGG~~ATGTTG~~TG~~CAGAGCTGAAGAGC  
 7422 --AA-CT---A-T---A-A---A-TA-C---C---  
 7434 TGGG...GAGAATGG...GGCTGGGCC~~ACCC~~AAGCAGGAGGCTGGG  
 7465 --T-CTCACT---T---AAA---T---T-TGAGTTT---A---AC  
 7475 ACGCTCT.GCTG~~TGGG~~CACAGGTCAG..GCTAATGT...TGGC  
 7515 C-A--G-G-ACA---G-G-A-A---AA-A---AT-AGCCGCTCCC---C-  
 7512 AGATGCAGCT~~CTTC~~CTGGA.CAGGCCAGGTGGTGGGCATT.CTCTCTCCA  
 7565 TA-GAT-C---AA-A---TA-T-A---CCA---A---AT-G---A---  
 7560 AGGTGTGCC~~CTGGG~~CATTACTGTTAA~~GACACT~~TCCGTACATCCAC  
 7615 ---CA---AA-A---G---CAG-C-G---T---T-CT---T  
 7610 CCCATCCTCC~~CAGGG~~CTAACAC...TGTGACATCTCTATTCCCCACCCCTC  
 7665 GTTGC---C-T---TA-A---GT---TAA-C---T---G---  
 7657 CCC~~TCCC~~CAGGGCAATAAA~~GACCAT~~GGAGGGGCTTG~~CAC~~TCTTGG  
 7708 G---A-G---T---GG---TAGCA---ACTC---T---CA  
 7707 CTGT~~CACCG~~GATGCC~~CAGC~~AAACTTAGATGTGAGAAAACCC~~CT~~CCCAT  
 7753 A---G-T---TA---TC---A---TC-G-GCC---T-C-A---GT---  
 7757 TCCATGGCGAAAACATCTCCTAGAAAAGCCATTACCC~~CT~~CATTAGGCATG  
 7800 ---T---A-C---C---T---TG---TT---GCAG---T  
 7777 GTTTGGGCT...CCAAAACACCTGACAGCCCC~~CT~~CCCTCTG  
 7782 ---CCA-C---AATGTAAGAGG---C-G-G-A-TGTT---T-GGAG---  
 7749 AGAGGCCGGAGAGT~~GCTGACTG~~TAGTGACCA.TTGCATGCCGGTGCAGCA  
 7893 ...T-T---C---T---AC---G---GC-ATA-TAGTT-TT---  
 7898 TCTGGAAGAGCTAGGCAGGGTGTCTGCC~~CCC~~CTCTGACTGAAGT~~CATG~~C  
 7941 G---C-A---A---ACA---AA-A-CC-T---TG-A---  
 7948 TCCCCTGTGCCAG~~CCC~~AGAGGCC~~GAG~~GCTATGGACAGCATT...GCCAG  
 7991 ---C---G---GG-T---A-C---T-G-AT-G-GCA---  
 7995 TAACACAGGCCACCC~~TGTG~~CAGAAGGGAGCTGGCTCCAGCCTGGAAACCT  
 8040 ...G---A---T---T---TT---T-A---TCAA---TC  
 8045 GTCTGAGGTGGAGAGGTGCACTTGGGCA~~CAGGGAGAG~~.GCCGGGACA  
 8079 A---A---CA---GCAA---G---A---A---AT---A---  
 8094 CACTTA...GCTGGAGATG~~TCT~~AAAAGCC~~CT~~TGATCGTATT~~CAC~~T  
 8128 G---C---GCTGG---GG---TG---C-G---  
 8139 TCAGTTTTGTGTTGGACAATTACTTTAGAAAATAAGTAGGT~~CG~~TT  
 8178 ---C---  
 8189 TAAAAACAAAATATTGATTGCTTTTG~~TAGT~~GTT~~CAGAA~~.AAAAGGT  
 8228 ---A-C---  
 8238 TCTTTG~~TG~~TATGCCAAATGACTGAAAGC~~ACTG~~ATATATTAAAACAAA  
 8276 ---  
 8288 AGGCAATTATTAAAGGAAATTG~~TAC~~CTTCA~~GTA~~AAAC~~CTG~~TCTGAATG  
 8326 ---  
 8338 TACCTGTATACGTTCAAAAACACCCCCCCCCACTGAATCC~~CTG~~TAACC  
 856 ...-A---C---  
 TATTTATTATATAAAGAGTTGCCTTATAAATTAA

gestrichelte Linie: Putativer Promotor

durchgezogene Linie: sequenzkonservierte Sequenz mit hoher Energie

1  
 human TTGCTGCAGATACTACTGACCAAGCTGTTGACCAGGCACCTCCCTCCGCCAAACCTT.....CCCCCATGTGGTCGT  
 schim  
 orang  
 makak  
 hamst  
 mouse  
 rat  
 kaeng

101  
 human TAGAGACAGAGCGACAGAGCAGTTGAGAGGACACTCCCGTTTCGGTGCATCAGTGCCTCGTCTACA...GCTCCCCAGCTCCCCC...ACCTCCCC  
 schim  
 orang  
 makak  
 hamst  
 mouse  
 rat  
 kanga

201  
 human ACTCCCAACCACGTT.GGGACAGGGAGGTGTGAGGCAGGAGACAGTT.GGATTCTTAGAGAAGA...TGGATATGACCAGTGGCTATGCCGTGCT  
 schim  
 orang  
 makak  
 hamst  
 mouse  
 rat  
 kanga

301  
 human GATCCCACCGTGGCTGGCTCAAGTCTGGCCCCACACCAGCCCCAATCCAAAAGTGGCAAGGACGCTTCACAGGACAGGAAAGTGGCACCTGTCTGCTGCC  
 schim  
 orang  
 makak  
 hamst  
 mouse  
 rat  
 kanga

401  
 human AGCTCTGGCATGGCTAGGAGGGGGAGTCCCTTAAGTACTGGGT.GTAGACTGGCTGAACCACAGGAGAGGATGGCCAGGGTGAGGTGGCATGGCC  
 schim  
 orang  
 makak  
 hamst  
 mouse  
 rat  
 kanga

501  
 human ATTCTCAAGGGACG.TCCTCCAACGGGTGGCGCTAGA...GGCCATGGAGGCAGTAGGACAAGGTGCAGGCAGGCTGGCTGGGTCAAGGCCGGCAG  
 schim  
 orang  
 makak  
 hamst  
 mouse  
 rat  
 kanga

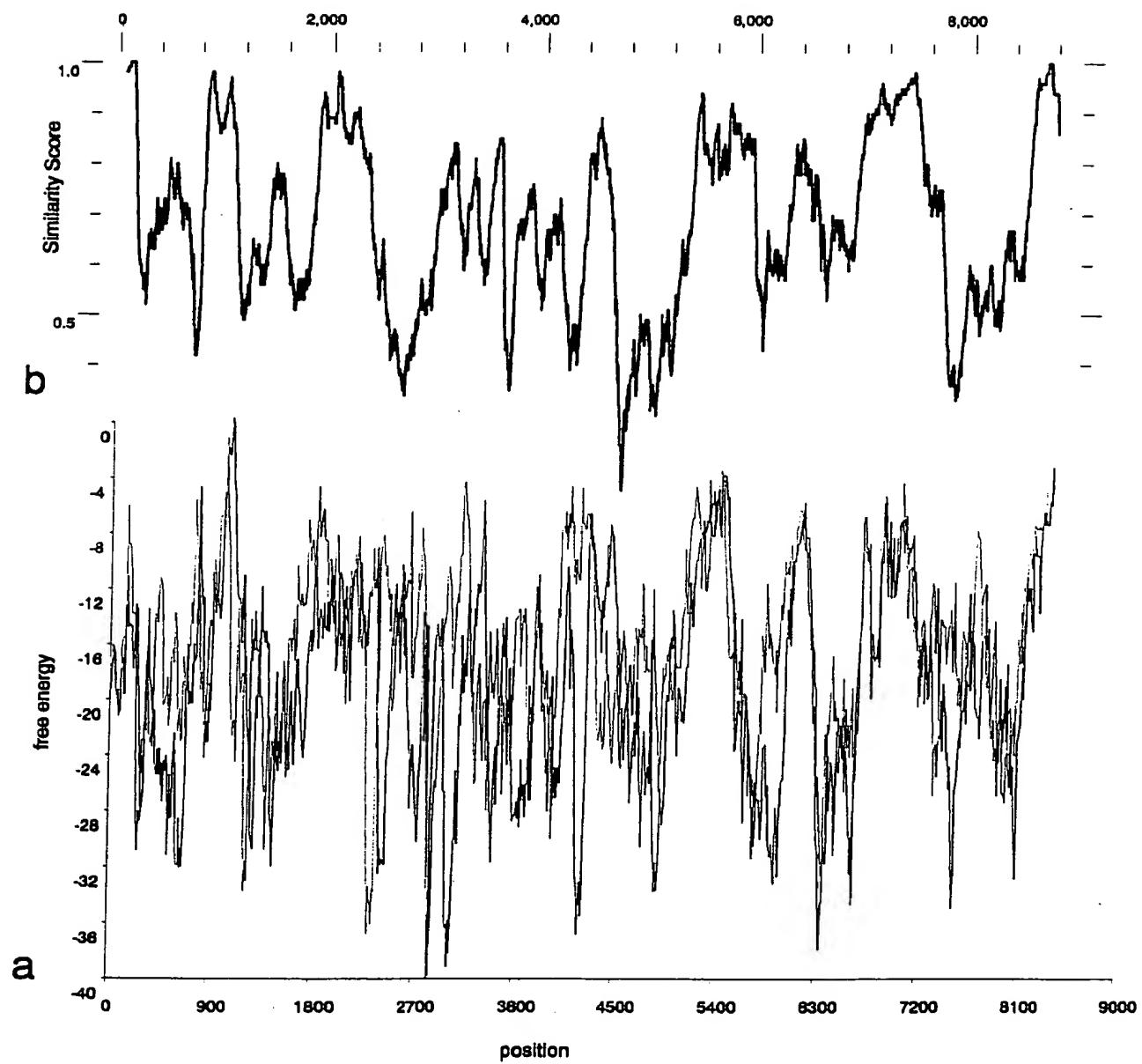
601  
 human AGCACAGCGGGGTGAGAGGGATTCTTAATCACTCAGAGCAGTCTGTGACT.....TAGTGGACAGGGGAGGGGGCAAAGGGGGAGGAGAAG  
 schim  
 orang  
 makak  
 hamst  
 mouse  
 rat  
 kanga

701  
 human AAAATGTTCTTCCAGTTACTTCCAATTCT...CCTTAGGGACAGCTAGAATTATTCGACTATTGAGTCTTCAT...GTTCCCACTTCAAAACAAA  
 schim  
 orang  
 makak  
 hamst  
 mouse  
 rat  
 kanga

801  
 human CAGATGC....TCTGAGAGCAAATGGCTTGAATTGGTGACATTAGTCCCTCAAGCCACCAAGATG....TGACAGTGGTGGAGAACTACCTGGATTT  
 schim  
 orang  
 makak  
 hamst  
 mouse  
 rat  
 kanga

901  
 human GTATATATACCTG  
 schim  
 orang  
 makak  
 hamst  
 mouse  
 rat  
 kanga

15/15



Dr. Coy - Abteilung Poustka 0840  
schwarz similarity 100 window  
blau hinlex 10 HUMAN  
grün minlex 10 MARS

